Norovirus evolution: Understanding and characterizing the emergence of novel strains in the population

Daniel Kelly

Supervisors: Prof Miren Iturriza-Gomara, Prof Nigel Cunliffe, Dr Alistair Darby

2017

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by Daniel Kelly

September 2017

Declaration

This thesis is the result of my own work and effort. The material contained in the thesis has not been presented, nor is currently being presented, either wholly or in part, for any other degree or other qualification.

Research in this thesis was carried out at the Institute of Infection and Global Health (University of Liverpool, UK).

MPS library normalisation, pooling, HiSeq (Illumina) sequencing and bioinformatic processing of the raw sequence data were performed by the Centre for Genomic Research (University of Liverpool, UK). Enzyme immunosorbent assays (EIA) to quantify total IgA and VLP specific IgA were performed by Anna Puławska-Czub. The human Norovirus (HuNoV) GII reference plasmid and GII.4 Farmington Hills recombinant baculovirus were kindly gifted to us by Dr David James Allen (The London School of Hygiene & Tropical Medicine, London). Partial sequence and epidemiological data, obtained on HuNoV outbreaks and seeding events (Chapter 5), was obtained from the Enteric Virus Unit (EVU) at Public Health England (PHE). Preparation of insect cell lysates for electron microscopy (EM) and EM imaging was performed by Dr Matthew Hannah (PHE).

Acknowledgements

I would like to thank my supervisors for their support and guidance in the past 4 years – Professor Miren-Iturriza Gomara, Professor Nigel Cunliffe and Dr Alistair Darby.

I would like to thank all past and present members of the Enteric virus research group, within the Institute of Infection and Global Health, and the Enteric virus unit, Public Health England, for their help throughout my studies. Most of all to Dr Khuzwayo Jere, Dr David James Allen and Winifred Dove for making me the molecular biologist I am today.

I would also like to thank my previous supervisors and technical staff, Dr Mal Horsburgh, Professor Alan McCarthy and Paul Loughnane, Lab H - Life Sciences, for providing me with an enjoyable undergraduate research experience, further encouraging me to move on to postgraduate study, and allowing me to use their equipment when ours was out of order.

I am extremely grateful to Public Health England for providing me with the opportunity and funding to perform this research.

I thank my parents and brothers for their unceasing encouragement, support and tolerance when I may have been intolerable. To my friends, and the friends I have made along the way, I have enjoyed our scientific discussions both in the lab and public house.

Abbreviations

A – Absorbance
AA – Amino acid
Ab - Antibody
AcMNPV – *Autographa californica Multicapsid Nucleopolyhedrovirus*
ATP – Adenine triphosphate
BMT – Bone marrow transplant
cAMP – Cyclic adenosine monophosphate
CAT – Chloramphenicol acetyl transferase
cDNA – copy DNA
CIC – Competitive internal control
CRFK – Crandell Reese Feline Kidney
CRT - Cyclic reversible termination
Ct – Cycle threshold
ddNTPs – dideoxynucleotides
dN – Rate of non-silent substitutions
DNA – Deoxyribonucleic acid
dNTPs – Deoxynucleotides
dS – Rate of silent substitutions
EIA – Enzyme immunoassay
EIC – Exogenous internal control
ELISA – Enzyme linked-immunosorbent assay
EM – Electron microscopy
ER – Endoplasmic reticulum
ES – External standard
FCV – Feline calicivirus
FH – Farmington Hills
FMDV – Food-and-Mouth Disease Virus
FUT2 - α(1,2)fucosyltransferase
GC – Gas chromatography
GE – Gastroenteritis
gEqs – Genomic equivalents
GI – Gastrointestinal
GADPH - Glyceraldehyde-3-phopshate dehydrogenase
Gn – Guanidium Isothiocyanate
Gnc – Gnotobiotic
HBGA – Histoblood group antigen
HCV – Hepatitis C Virus
HIV – Human Immunodeficiency Virus
HuNoV – Human Norovirus

IAPV – Israeli Acute Paralysis Virus
IAV – Influenza A Virus
IC – Internal process control
ICNV – International committee on nomenclature of viruses
ICTV – International committee on taxonomy of viruses
ID_{50} – Infectious dose 50
IEM – Immunoelectron microscopy
IgA – Immunoglobulin A
IgG – Immunoglobulin G
INF – Interferon
ITOL – Interactive Tree of Life
LAMP – Loop mediated isothermal amplification
MAb – Monoclonal antibody
MAPREC – Mutant Analysis by Polymerase Chain Reaction and Restriction Enzyme Cleavage
MBs – Magnetic beads
MNV - Murine Norovirus
MPL – Monophosphoryl Lipid A
MPS – Massively parallel sequencing
mRNA – Messenger RNA
MS – Magnetic separation
MV – Measles virus
NASBA – Nucleic acid sequence-based amplification
NIC – Non-competitive internal control
NLV – Norwalk-Like Virus
NO – New Orleans
NoV – Norovirus
NV – Norwalk Virus
OAS - Original antigenic sin
OPV – Oral poliovirus vaccine
ORF – Open reading frame
PAbs – Polyclonal antibodies
PBS – Phosphate buffered saline
PCR – Polymerase chain reaction
PDB – Protein data bank
pfu – Plaque forming units
PGM – Porcine gastric mucin
PGM-MB – Porcine gastric mucin-Magnetic bead
pH – Potential of hydrogen
PHE – Public Health England
PPE – Personal protective equipment
ppm – Parts-per million
Q – Quality score
qPCR – Quantitative PCR

RBC – Red blood cell
RC – Replication complex
RdRp – RNA dependent RNA polymerase
RHDV – Rabbit Haemorrhagic Disease Virus
RNA – ribonucleic acid
RoV – Rotavirus
RSV – Respiratory Syncitial Virus
RT – Reverse transcription
RT-PCR – Reverse transcription-polymerase chain reaction
RTS – Real-time sequencing
SBL – Sequencing by ligation
SD – Standard deviation
SDW – Sterile distilled water
SNA – Single nucleotide addition
ssRNA – Single stranded ribonucleic acid
SY – Sydney
TMB – 3',3',5',5'-tetramethylbenzidine liquid substrate
UTR – Untranslated region
VDPV – Vaccine-derived poliovirus
VLP – Virus-Like particle

Abstract

Human noroviruses (HuNoVs) are distributed globally, affect all age groups and place a significant burden upon health services. The diversity of this RNA virus is thought to play a significant role in the persistence of HuNoVs as the main cause of non-bacterial gastroenteritis globally. Molecular diagnostics have been critical for understanding the epidemiology of outbreaks and sporadic cases, and to design and implement effective intervention strategies and disease control measures. Immunocompromised individuals are widely considered to be a reservoir for epidemic variants of HuNoV and whilst there are studies investigating the emergence of novel strains in an immunocompetent general population, reports at the the individual level are scarce.

Three separate methodologies were developed to characterise HuNoV persistence in acute convalescent and chronic infection. First, a standardised quantification method to accurately quantify the most prevalent HuNoV genogroup. Second, a PGM-MB capture method to select HuNoV prior to massively parallel sequencing (MPS). Third, an assay to measure host specific coproantibody responses to three epidemic variants from different epochs.

Quantification of longitudinal samples from individuals with acute or chronic HuNoV infection showed the virus distribution was homogenous in stool and an RNA external standard, in contrast to DNA, did not underestimate virus titre. HuNoV PGM-MB capture meant near complete viral genomes could be recovered at variable mean coverage. A bioinformatics pipeline demonstrated over the course

of chronic infection allele frequencies were much more variable. In acute infection, minor alleles were present at a much lower frequency, but potential immune escape mutants were present. Immune escape mutants existed as minority variants or conserved mutations in the consensus sequence, and were in the presence of HuNoV specific-coproantibody, which were mapped to the protein surface.

In HuNoV chronic infection, immune pressure is variable or non-existent, and therefore epidemic variants could emerge over long periods of infection by random chance. However, under immune pressure exerted by coproantibodies, escape variants may be seen. In three individuals, acute HuNoV symptomatic infection occurred despite the presence of specific secretory Ab responses to the VLP classed as the closest phylogenetic relative. The closest relative (Sydney 2012), differed at two amino acids, one of which has been previously described (A340T) as belonging to an epitope, and another which can be classed as having a potential role in immune escape (A323T). A single individual with acute HuNoV infection established a more prominent response to an earlier strain of HuNoV, rather than two contemporary strains, which proposes a role for Original Antigenic Sin (OAS) or Antigenic Seniority in the secretory Ab immunity.

Finally, the use of MPS in outbreak tracking was assessed and compared to the currently used amplicon and Sanger based method. Overall both methods showed significant correlation. However, MPS provided greater depth and the ability to identify variants among samples within an outbreak that represented consensus changes in one or more samples from the same outbreak. This meant that the MPS data would have been able to link all the samples into a single outbreak or

transmission network, where the current Sanger sequencing may not have been able to link them all.

Table of Contents

Figures

Tables

1 Introduction

1.1 Historical background

Zahorsky was the first to describe “winter vomiting disease”; an enteric illness that was predominantly observed between September and March (Zahorsky 1929). Thereafter, a significant proportion of sporadic and epidemic gastroenteritis (GE) was referred to as non-bacterial with no discernible cause (Adler, J. L. 1969; Gordon, I., H. S. Ingraham 1947; Hodges et al. 1956). In subsequent studies the stool filtrate from symptomatic individuals was used as an inoculum to infect asymptomatic volunteers, and from this the transmission route of non-bacterial GE was found to be faecal-oral (Gordon, I., H. S. Ingraham 1947; W.S. et al. 1953). Serial propagation of the inoculum between volunteers induced a similar incubation period, clinical course and symptom duration to that experienced during outbreaks, implying a viral agent or agents to be responsible (Adler, J. L. 1969; Clarke et al. 1972; Dolin et al. 1971; Gordon, I., H. S. Ingraham 1947).

In 1972, Kapikian was the first to observe a causative agent of non-bacterial GE by using serum from convalescent individuals to capture the virus in stool, from the acute phase of infection, prior to electron microscopy (EM), also known as immunnoelectron microscopy (IEM) (Kapikian et al. 1972). The virus identified in the study ranged in diameter from 27 to 32 nm and was aggregated by the serum of naturally infected individuals or volunteers, to show a causal link between the agent and clinical disease (Kapikian et al. 1972). Moreover, these experiments determined antigenic relatedness between the agent from an elementary school outbreak (Norwalk, OH, USA), and the virus found during the reproduced illness in

volunteers (Kapikian et al. 1972). The virus present in these investigations was termed the Norwalk virus (NV) agent (Kapikian 2000).

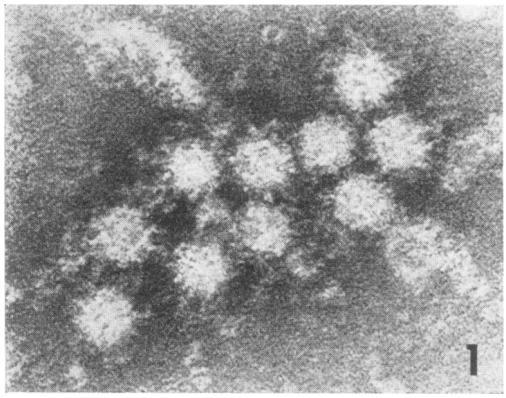


Figure 1.1. IEM image of aggregated NV (Kapikian et al. 1972).

In the years following, a methodology using convalescent host sera was developed to elucidate virus characteristics and disease epidemiology (Greenberg and Kapikian 1978). Seroprevalence studies were undertaken which demonstrated NV to be prevalent, in all age groups, across multiple continents with an exposure rate of 70% by adulthood (Blacklow et al. 1979; Greenberg et al. 1979).

Over a decade later Jiang and colleagues cloned and sequenced the NV genome, which was critical for the development of contemporary diagnostic tools (Jiang et al. 1990). Consequently, the complete NV nucleic acid sequence led to the design of an assay to detect the virus nucleic acid by reverse transcription polymerase chain reaction (RT-PCR) and numerous recombinant expression systems for the viral proteins (Baric et al. 2002; Jiang et al. 1992a; Jiang et al. 1992b). Recombinant expression of the virus capsid protein has provided a limitless amount of antigen for serological assays, and together with molecular techniques have been pivotal in the assessment of virus burden, diversity and spread.

Despite advances in the laboratory and a better understanding of disease epidemiology, much about the nature of the virus remains unknown. The lack of a reproducible virus culture system or a suitable small animal model has been a major barrier in understanding the virus stability, host cell-virus interactions and disease development (Duizer 2004; Tan and Jiang 2010). However, the recent developments of two different *in vitro* culture models will provide new avenues to research the virus, that were not possible previously (Ettayebi et al. 2016; Jones et al. 2014).

1.2 Taxonomy

There are two main virus grouping systems that have been developed, the hierarchical system in 1962 and the Baltimore classification in 1971 (Baltimore 1971; LWOFF et al. 1962). The hierarchical system arranges a virus by the composition of its genetic material, capsid structure or the presence of an envelope. On the other hand, the Baltimore classification differentiates a virus by genomic material composition only, due to its influence on the mechanism of mRNA synthesis and facilitation of protein expression. In 1966, the international committee on taxonomy of viruses (ICTV), then the international committee on nomenclature of viruses (ICNV), was formed and accepted the hierarchical system to standardise virus classification. The committee assigns newly discovered viruses to taxons and maintains current groupings, under the definition 'A virus species is a polythetic class of viruses constituting a replicating lineage and occupying a particular ecological niche' (Vanregenmortel 1989).

Several antigenically distinct viruses similar in morphology to NV were isolated from outbreaks matching the epidemiological pattern of the Norwalk agent, and thus termed the Norwalk-Like viruses (NLVs) (Dolin et al. 1982; Thornhill et al. 1977; Wyatt et al. 1974). The newly discovered NLVs were grouped with the prototype virus by serology, or IEM and enzyme linked-immunosorbent assay (ELISA), or partial genomic sequencing, in the absence of a virus culture system. However, early grouping systems lacked concordance due to genome sequencing of short regions with relative rates of mutation, virus recombination and reagent variability in serological assays (Cubitt et al. 1987; Green et al. 1993; Jiang et al. 1995; Lew, J F, Kapkian, A Z, Jiang, X, Estes, M K, Green 1994; Treanor et al. 1993).

NV, and the NLVs, were classified into the *Calicivirdae* family, defined by the ICTV in 1981, and its name derived from the Latin *calyx,* to mean “cup” or “goblet” (Green 2007). The name, *Caliciviridae*, describes the cup-shaped depressions that can be observed on the surface of the family by EM. In 2002, the ICTV grouped these viruses under the Genus Norovirus (NoV), which currently exists alongside Lagovirus, Nebovirus, Sapovirus and Vesivirus (Carstens and Ball 2009; Green et al. 2000; Mayo 2002; Pringle 1999). Noroviruses are known to infect both animals and humans, whilst Lagoviruses, Neboviruses and Vesiviruses infect animals and Sapoviruses appear to be restricted to humans (Green 2007). In the family, Norovirus genome structure is most like Vesivirus, with three open reading frames (ORFs) rather than two (Figure 1.2).

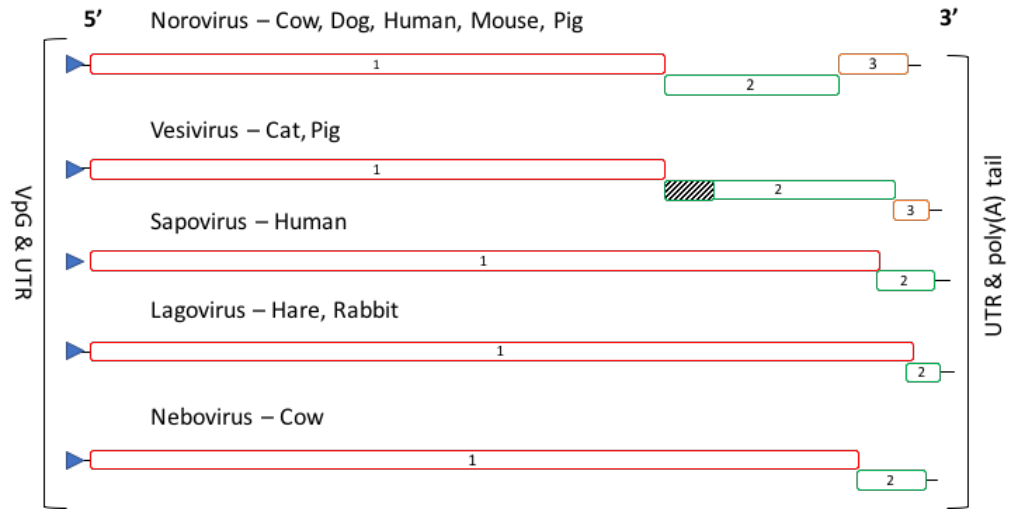


Figure 1.2 The genome ORF structures of the *Caliciviridae*. The shaded region of Vesvivirus corresponds to a leader sequence of the capsid precursor protein. Adapted from Field's Virology (Green 2007), Accession numbers: Norovirus – M87661, Vesivirus - M86379, Sapovirus – X86560, Lagovirus – M67473, Nebovirus – JX018212.

A definitive NoV classification does not currently exist, although interim grouping systems have been proposed based on capsid sequence data (Ando et al. 2000; Fankhauser et al. 2000; Kageyama et al. 2004; Vinjé et al. 2000). These systems have applied genetic distance comparisons to variants to define genetic clusters. The most recent published interim-grouping scheme defines 7 genogroups and 41 genotypes (Kroneman et al. 2013; Vinje 2015). In the literature, a genogroup is expressed with roman numerals (9 GI, 22 GII, 3 GIII, 2 GIV, 2 GV, 2 GVI and 1 GVII), whilst an arithmetic value is appended to the genogroup to describe the genotype. For example, a virus in the NoV Genus labelled GII.4, would mean genogroup 2 and genotype 4. Human norovirus (HuNoV) infection is predominantly caused by GI, GII and to a much lesser extent GIV (Figure 1.3).

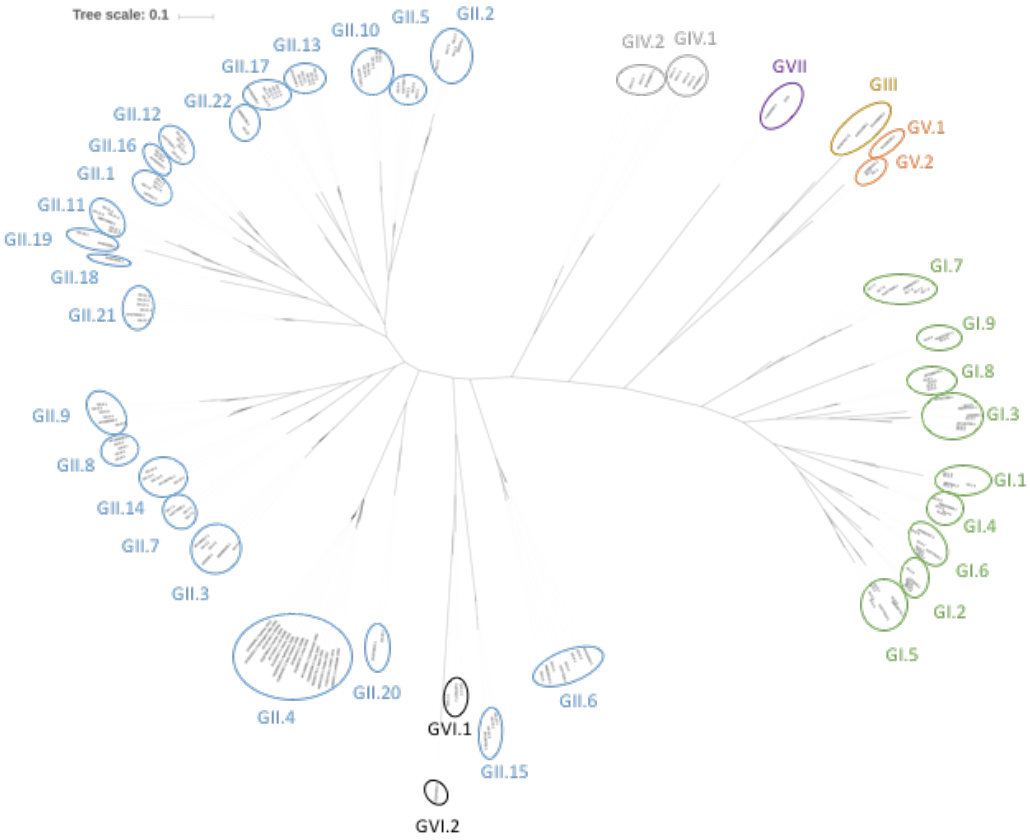


Figure 1.3 An unrooted Maximum-Likelihood Tree of the NoV Genus (Genogroups I to VII) based on partial and complete nucleic acid sequences of open reading frame (ORF) 2 (Green=GI, Blue=GII, Yellow=GIII, Grey=GIV, Orange=GV, Black=GVI and Purple=GVII).

1.3 Human norovirus diversity

HuNoVs are highly diverse, and there are two main mechanisms that are responsible for the promotion of this diversity. Firstly, drift is the accumulation of diversity through random mutations, and secondly recombination, which is reported to occur when two or more viruses from the same genogroup enter a cell to form a chimeric progeny of viruses (White 2014).

1.3.1 Drift

Drift is a mechanism by which RNA viruses randomly mutate at a rate of 1 substitution in every 1,000 to 100,000 bases, and this is largely due to the RdRp

(RNA dependent RNA polymerase) lacking a proof reading mechanism (Domingo et al. 1997). The RdRp incorporates nucleotides at low fidelity and high velocity, and these two properties act in tandem to form quasispecies, or a population that exists as a mutant swarm rather than separate genetic entities (Domingo et al. 2012; Eigen 1971).

The principle of quasispecies was first introduced in 1971 by Eigen from investigating the Qβ replicase system. The use of "evolution in the test tube" allowed for the isolation of a variety of phage descendants with subtle changes to RNA genome structure under different conditions (Eigen 1971). This term was further defined as an organised population of species with a defined probability distribution dependent upon selection (Eigen and Schuster 1977). Experimentally each species is not thought to exist in a steady state, with competing mutants constantly moving in and out of existence under changing environmental selection (Domingo and Holland 1997).

Individual swarms of related viruses cooperate with each other allowing for less fit mutants to become dominant, assuming closely related viruses within the swarm are better adapted in comparison to other swarms that may contain a more fit virus (Domingo, Martin et al. 2006). Populations of these RNA viruses are able to rapidly explore sequence space expressing most of the viable viruses possible within each replication cycle including every point mutation and many double mutations (Vignuzzi, Stone et al. 2005). This can lead to alterations in virus cell tropism, antiviral nullification and antibody evasion (Domingo, Sheldon et al. 2012).

The role of quasispecies in driving evolution over the course of infection is well studied in Human immunodeficiency virus (HIV) and members of the *Flaviviridae* for promoting evolution over the course of infection. An example of the impact of the evolution of these populations would be the use of various chemokine receptors instead of C-C chemokine receptor type 5 (CCR5) only by HIV, to prevent competition with chemokines and allow for a broader tissue spread (Connor, Sheridan et al. 1997). Both Hepatitis C virus (HCV) and HIV infections can require long term antiviral suppression, potentially selecting for and establishing drug resistant mutants within virus populations (Margeridon-Thermet and Shafer 2010). In addition antibody selective pressures were found to correlate with increased mutation rates of HCV quasispecies, and this was reduced in patients unable to mount an adaptive immune response (Di Lorenzo, Angus et al. 2011).

Drift is believed to be a major mechanism driving diversity and emergence of novel HuNoVs and the emergence of pandemic strains. In the late 1990's the first pandemic HuNoV was detected, belonging to GII.4, and since then this genotype has predominated globally for the last two decades (Kroneman et al. 2008b; Noel et al. 1999). However, in the GII.4 genetic cluster, periodic strain replacement has occurred every 2-3 years (Figure 1.4). This phenomenon, termed epochal evolution, and first described in Influenza A viruses, is believed to be driven primarily via selective pressure exerted by herd immunity (Allen et al. 2008; Siebenga et al. 2007). Such mutations therefore will lead to the emergence of antibody escape mutants in waves, which will spread through the population from time to time giving rise to pandemics.

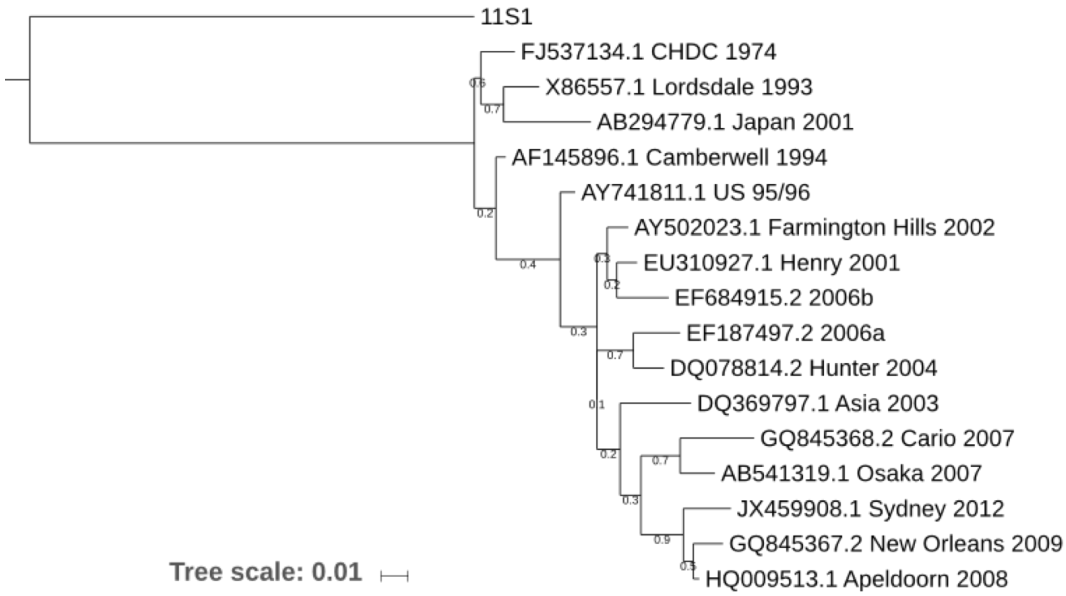


Figure 1.4 A rooted Maximum-Likelihood tree (1000 Bootstrap replicates) of GII.4 epidemic strains based on complete nucleic acid sequences of the major capsid protein adapted from (Eden et al. 2013). 11S1 corresponds to a GII.7 virus sequenced in this study.

1.3.2 Recombination

Recombination occurs when two or more different viruses infect the same cell and exchange genetic material, and thus, can promote diversity in a virus population (Pérez-Losada et al. 2015). In viruses, recombination can expand the host range, modify tissue tropism, alter virulence, allow evasion from the host immune response or induce antiviral resistance (Pérez-Losada et al. 2015).

HuNoV recombination is postulated to occurs in a similar manner to HCV and Poliovirus (Simmonds 2006), via the copy choice model, or template switching, and it has been best and more frequently described at the ORF1+ORF2 region, where the highest degree of conservation exists (Bull et al. 2005; Bull et al. 2007; Eden et al. 2013; Lindesmith et al. 2008; Rohayem et al. 2005; Teunis et al. 2008; Waters et al. 2007). Recombination occurs among HuNoV of the same genogroup, and despite

some potential inter-genogroup recombination having been described (Nayak et al. 2008), the evidence is to date not sufficiently robust. Template switching is proposed to occur when the RdRp encounters a lack of contiguous sequence, through an RNA structure or a change in genome composition, at which point the polymerase jumps to a different template (Bull et al. 2005; Rohayem et al. 2005). To accurately detect a recombination event, complete sequence data is required from the ORF, or the two adjacent ORFs if the breakpoint is in a junction.

1.3.3 Diversity and persistence of GII.4 strains in the human population

It is likely that the persistence of GII.4 strains in the human population is a consequence of both mechanisms, drift and recombination. Both mechanisms may lead to, as mentioned previously, the emergence of antibody escape mutants, but also the emergence of strains with a broader range of host binding specificities or differences in replication rates. Collectively, these factors are potentially responsible for the successful establishment of GII.4 over other genotypes. The two most contemporary replacements, New Orleans 2009 and Sydney 2012, are recombinants of older strains (2006a/Apeldoorn 2008 & Osaka 2007/Apeldoorn 2008, respectively) (Eden et al. 2013). Therefore, drift and recombination contribute to the selection of GII.4 strains and, in an analogous manner to Influenza A strains, appear to enable strains to undergo epochal evolution and escape from herd immunity (Lam et al. 2012; Lindesmith et al. 2008; Siebenga et al. 2007).

1.4 Epidemiology

1.4.1 Burden

HuNoV is an aetiological agent of sporadic and epidemic GE in all age groups worldwide (Ahmed et al. 2014). It is recognised as the most common cause of sporadic GE, in the community, and GE outbreaks especially in semi-closed environments such as health and social care institutions, schools and nurseries, holiday camps and cruise ships (Hall et al. 2013; Inouye et al. 2000; Lopman et al. 2003; Meakins et al. 2003; Tam et al. 2012). However, data from developing countries, with high mortality rates due to diarrhoea, are underrepresented in the literature (Ahmed et al. 2014). Moreover, a recent study identified a lack of data on HuNoV illness burden estimates in age groups 5 years and over, especially in low or middle income countries (Bartsch et al. 2016). Nevertheless, the limited data available has estimated HuNoVs to be a significant cause of mortality in children less than 5 years old, mostly in developing countries (Bartsch et al. 2016; Lanata et al. 2013; Patel et al. 2008). In developed countries, the virus places the greatest burden on paediatric populations, whilst the most severe clinical outcomes tend to occur in the elderly (Lindsay et al. 2015). Since the implementation of a rotavirus vaccine, HuNoV has become the leading cause of paediatric hospitalisation due to GE (Payne et al. 2013).

The high prevalence of HuNoV in children or the increased mortality rate in the elderly show that, although all age groups are susceptible to infection, these individuals are under an increased risk of exposure or severe illness, respectively. Molecular epidemiology of HuNoVs has made it possible to associate risk of

genotype infection to an age group, setting or transmission route. Globally, GII HuNoVs are more prevalent than GI in the community and outbreaks, whereas GIV infections are rare (Sakon et al. 2015; Vega et al. 2014a). Of the existing GII viruses, GII.4 is the most common cause of HuNoV related sporadic GE in children and outbreaks in older age groups (> 65 years old) (Hoa Tran et al. 2013; Vega et al. 2014a). Interestingly, of the HuNoV sporadic illness burden in paediatric populations, GII.3 has been found to be highly prevalent, however it is significantly less prevalent in individuals over 60 years of age (Franck et al. 2014). Of HuNoVs, GII.4 is a common cause of foodborne illness and the most common cause of sporadic illness in children, but it is more regularly associated with person-to-person transmission and outbreaks in healthcare environments than foodborne outbreaks or sporadic illness (Franck et al. 2014; Sakon et al. 2015; Vega et al. 2014a). The less prevalent GI viruses have been reported to be more significantly associated with foodborne outbreaks than GII.4 strains (Franck et al. 2014; Vega et xal. 2014a). GI viruses are often detected in spring or summer months in contrast to the autumn/winter seasonality that is driven primarily by GII.4 HuNoVs (Franck et al. 2014; Vega et al. 2014a).

The predominance of GII.4, over GII.3 and other genotypes, in institutional outbreaks and the community is likely due to a combination of factors: a higher mutation rate and thus greater ability to evade the herd immunity, increased virulence, a broad host receptor binding specificity, greater stability and survival in the environment and a lower virus dose to cause disease (Bull and White 2011).

It has been demonstrated that the GII.3 polymerase incorporates nucleoside triphosphates more slowly and has a higher fidelity in comparison to the corresponding GII.4 enzyme (Bull et al. 2010). However, a different study attributed GII.4 predominance to an increase in the fixation of amino acid substitutions rather than nucleotide mutation rate (Boon et al. 2011). Therefore, a low GII.3 prevalence could be a consequence of a relative inability to evade herd immunity, the virus being maintained by the replenishment of a susceptible population, through the birth of immunologically naïve children (Franck et al. 2014). Secondly, GII.4 HuNoVs have been associated with more severe illness both in children (prolonged vomiting and diarrhoea) and in the elderly (prolonged diarrhoea and higher levels of viral shedding) (Huhti et al. 2011; Lee et al. 2007). Thirdly, GII.4 can bind a broader range of Histo-blood group antigens (HBGAs), a known host attachment receptor, and in theory creating a larger pool (~80% of the population) of susceptible individuals (Boon et al. 2011; Huang et al. 2005; Yang et al. 2010). Fourthly, during infection, an immunocompetent individual can shed stable virus 3 to 14 hours prior to the onset of clinical disease, and virus can remain detectable up to 4 to 8 weeks after the resolution of symptoms (Atmar et al. 2008; Duizer et al. 2004; Rockx et al. 2002). Lastly, several human challenge studies have found the infectious dose of HuNoVs to be low, but there are variations between the studies. For a disaggregated GI.1 virus, the infectious dose 50 (ID_{50}; which is the dose required to infect 50% of a study population) for human volunteers ranged from 18 to 1015 genomic equivalents (gEqs) (Teunis et al. 2008). A more recent study of aggregated virus estimated a higher ID_{50} of 2800 gEqs, when all individuals of different genetic susceptibilities were included (Atmar et al. 2014). However, since GII is distinct from GI it would be

expected that the infectious dose is different, and possibly lower given the differences in prevalence. A GII.4 challenge study infected 70% of individuals at a dose of 50,000 gEqs, implying the ID_{50} is below this value (Frenck et al. 2012). Whether HuNoV during these latter stages of recovery from illness (or in the convalescent phase) is infectious and can lead to disease transmission is not known.

1.4.2 Transmission

HuNoV is transmitted by the faecal-oral route with numerous transmission routes contributing to the overall burden of HuNoV (Figure 1.5). The virus can be spread indirectly, if the environment or a fomite is contaminated (environmental transmission), or directly, by contact with a symptomatic individual (person-to-person transmission). A close genetic relatedness between human and porcine NoVs within GII, suggests the potential for interspecies transmission. However, zoonotic or anthroponotic transmission has not yet been proven despite the ability of HuNoV to infect pigs in experimental conditions (Cheetham et al. 2006). Therefore, the overall burden of HuNoV disease is known to be sustained by environmental or person-to-person transmission.

1.4.2.1 Environmental transmission

Environmental transmission is sustained by the introduction of HuNoV to a water source or food source (usually by sewage), fomite (i.e healthcare equipment) or during the preparation of complex foods (cross-contamination). A waterborne infection can originate from drinking or recreational water, whereas a foodborne infection can occur when fruit, vegetables (irrigation) or shellfish (bioaccumulation) encounter contaminated water (Mathijs et al. 2012). Furthermore, complex foods

can be contaminated during the preparation process if a food handler is shedding virus, either from asymptomatic infection or during recovery from illness (Parashar et al. 1998).

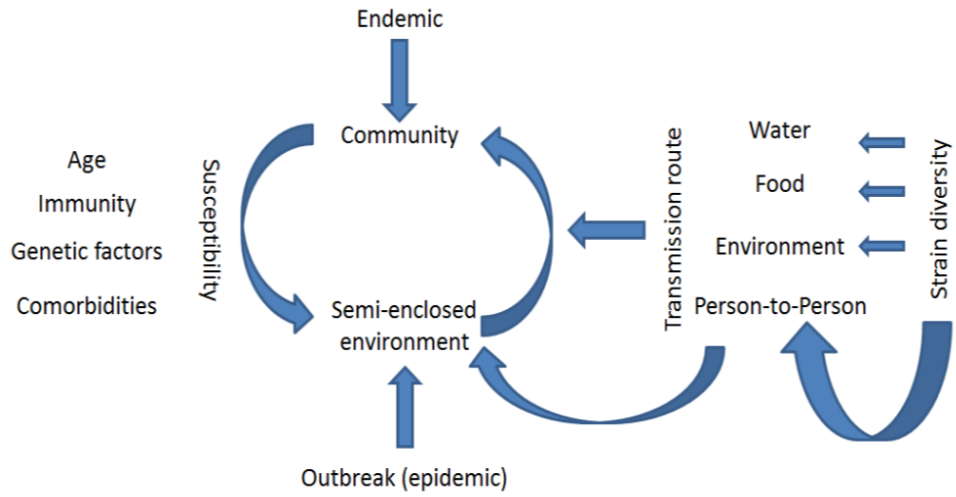


Figure 1.5 Transmission route and susceptibility factors influence how HuNoV cycles between the community (sporadic illness) and semi-closed environments (outbreaks).

In the literature, although GII NoV is the most common, the distribution of GI and GII NoVs responsible for waterborne outbreaks is highly heterogeneous (Kroneman et al. 2008b; Maunula et al. 2005). In Sweden, GI outbreaks from a water source were predominant, and this is consistent with a systematic review of published data which concluded waterborne NoV was more likely to be GI than GII (Lysen et al. 2009; Matthews et al. 2012). The observed higher frequency of GI waterborne outbreaks in comparison to other transmission routes, is proposed to reflect the stability of GI NoV in water (Ngazoa et al. 2008; Seitz et al. 2011).

In contrast, NoV foodborne outbreaks are more homogenous in genotype (for both GI & GII) distribution. This was highlighted in two studies, undertaken over similar time periods, in Europe and the US which found GII (85% and 83 %, respectively) to

be more abundant than GI (15% and 17%, respectively) (Hall et al. 2012; Kroneman et al. 2008a). In the US, foodborne outbreaks most commonly occur in a restaurant setting and infected food handlers were implicated as the main source of contamination (Hall et al. 2012). GII NoV has been identified to have greater stability on surfaces (Liu et al. 2010), and to be shed at viral loads 100-fold higher than GI (Chan et al. 2006) and these factors are likely to contribute to its prevalence in foodborne illness. Interestingly, GI.1, and to a lesser extent GII.3, NoVs were found to bio-accumulate in oysters, by specific binding of the digestive tissues, whereas GII.4 was noted to be unstable and bind transiently (Maalouf et al. 2010).

NoV water and foodborne transmission is more likely to be associated with mixed genotype outbreaks (GI and GII) and higher attack rates in comparison to other transmission routes (Matthews et al. 2012). Those outbreaks linked to multiple genotypes tend to be linked to the introduction of sewage contaminants from increased rainfall, and flooding, or inadequate wastewater plant treatment (Le Guyader et al. 2006; Hewitt et al. 2011). On the other hand, asymptomatic or post-symptomatic, food handler induced outbreaks would be associated with an identical HuNoV strain (Xerry et al. 2009).

1.4.2.2 Person-to-person transmission

In a semi-closed environment (e.g hospital, nursing home, day-care centre), an outbreak is propagated primarily by person-to-person transmission. Transmission between hosts occurs directly, through contact with aerosolized vomitus (Green et al. 1998), or indirectly via the spread of vomitus or faecal material onto hands, surfaces or equipment.

In person-to-person outbreaks, GII NoVs are the most common, and a large proportion of these infections are caused by the GII.4 pandemic strains (Kroneman et al. 2008b). Moreover, the GII, and GII.4, viruses are more likely to be detected in a healthcare or leisure setting and are seasonal, with a tendency of more GII outbreaks to be detected in the winter (Kroneman et al. 2008b; Matthews et al. 2012).

1.4.2.3 Zoonotic transmission

It is well established that members of the NoV Genus are aetiological agents in pigs (GII), cattle (GIII), dogs (GVI) and mice (GV). More distantly related members would need to jump the species barrier, although an increase in seroprevalence to GIII and GVI NoVs has been detected in veterinarians (Mesquita et al. 2013; Widdowson et al. 2005). HuNoVs (GII) have been proven to replicate and cause disease in gnotobiotic (Gnc) pigs (Cheetham et al. 2006), and viral RNA has been detected in pigs, cattle and retail meat, which would suggest an indirect transmission route through the food chain (Mattison et al. 2007).

1.5 Norovirus structure and composition

NoV is a non-enveloped virus of 27-32 nm in diameter and icosahedral symmetry. Due to its appearance by EM, NoVs were also known as small round structured viruses. The HuNoV virion comprises a positive sense single stranded RNA (ssRNA) genome of ~7.5-7.7 kb in length, flanked at the 5’ end by a protein structure (Vpg), analogous to the eukaryotic messenger RNA (mRNA) cap structure (Sachs et al. 1997), and at the 3’ end by a poly-(A) tail (Figure 1.6) (Green 2007).

1.5.1 Norovirus genome

The HuNoV genome comprises three ORFs, whilst a fourth ORF is present in murine norovirus (MNV) (Figure 1.6). ORF1 encodes the non-structural proteins p48, an NTPase, p22, VPg, $3CL^{pro}$ and the RdRp which are post-translationally cleaved (Figure 1.6 and Table 1.1). Unlike ORF1, ORF2 and 3 are expressed separately, and encode the major, VP1, and minor structural proteins, VP2, respectively (Figure 1.6 and Table 1.1). In MNV the fourth ORF encodes a virulence factor (VF1), which thus far has not been detected in other NoVs (Figure 1.6 and Table 1.1).

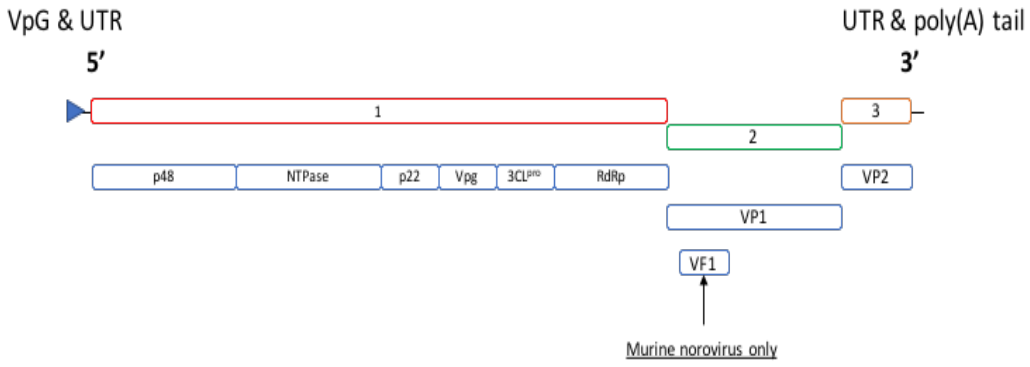


Figure 1.6 The ORF structure of a NoV genome, and the proteins expressed.

Table 1.1 The functions of proteins expressed by NoVs.

MNV	HuNoV	Function
NS1/2	P48 (N-term)	Replication complex formation, contributes to persistence in MNV infections
NS3	NTPase (2C-like)	RNA helicase/NTPase
NS4	P22 (3A-like)	Replication complex formation
NS5	VPg	Genome-linked protein involved in translation and replication
NS6	Pro (3C-like)	Protease
NS7	Pol/3Dpol	RdRP
VP1	VP1	Major capsid protein
VP2	VP2	Minor capsid protein
VF1	No equivalent	Virulence factor

Vpg

The Vpg protein structure is covalently attached to the 5' end of the NoV genome; it is analogous to the eukaryotic mRNA cap structure, and has multiple, integral, roles for virus infection (Herbert, Brierley et al. 1997). Firstly, in MNV It was found to bind the eIF3d/eIF4E sub units of the eIF3 complex involved in ribosome translation of mRNA and inhibited the production of host capped RNA (Daughenbaugh, Fraser et al. 2003). These interactions were found to be highly conserved been HuNoV and MNV when analysing VPg AA sequence and comparing protein binding patterns in a cell free system (Daughenbaugh, Wobus et al. 2006). Secondly, Vpg undergoes nucleotidylylation in the presence of the RdRP or the 3CLpro-RdRp intermediate, to act as a primer for NoV replication (Belliot et al. 2008). A Tyrosine at position 27 was found to be the site of nucleotidylylation, whereas groups of charged residues at the N terminus were found to be of functional importance (Medvedev et al. 2017). Thirdly, although no experimental evidence exists, Vpg protein priming has been postulated as a mechanism to avoid the formation of 5' triphosphorylated RNA, which would otherwise be detected by innate immune mechanisms in the host cell (Goodfellow 2011).

p48

Investigations to elucidate the role of p48 used protein expressing plasmids and cell target VAP-A in conjunction with yeast hybrids (Ettayebi and Hardy 2003). Although p48 function is unknown, *in vitro* it has been found at the golgi apparatus, and it is believed to promote both membrane rearrangement and protein trafficking within the host cell (Donaldson, Lindesmith et al. 2008).

NTPase

An NTPase was identified in early studies by comparing feline calicivirus (FCV) and picornavirus sequence alignments (Neill 1990). Similar enzymes in both genera were thought to be putative SF3 helicases, but p41 has been found to be unable to unwind a DNA-RNA heteroduplex and therefore is postulated to be a structural component in organising replication (Pfister and Wimmer 2001).

p22

The NoV p22 is similar to a traditional di-acidic endoplasmic reticulum (ER) export signal and antagonises protein secretion via golgi apparatus disassembly to inhibit the cell immune response (Sharp, Guix et al. 2010). This protein has been labelled as the most diverse protein amongst norovirus strains, yet it contains a conserved membrane association domain and a YXΦESDG motif (X is any AA and Φ can be M, I or L) found only in viruses with human tropism (Sharp, Guix et al. 2010).

3CLpro

In vitro translation of ORF1 demonstrated the "3C-like" protease (3CLpro) to be responsible for self-cleavage of the virus polyprotein product (Belliot et al. 2003). The roles of enzyme AAs His30 and Cys139 were identified as proteolytic activity with Glu54 responsible for substrate specificity (Someya and Takeda 2009). This protease can also inhibit the translation of cellular poly-adenylated mRNAs to host proteins through the cleavage of poly-(A) binding protein (Kuyumcu-Martinez, Belliot et al. 2004).

RdRp

The RdRp protein utilises the parental viral RNA for *de novo* and Vpg-dependent synthesis of negative sense genomic and subgenomic RNAs (Rohayem et al. 2006). Subsequent generations of viral positive sense RNA are formed by 3' Vpg protein priming, where positive sense subgenomic RNAs are generated by premature termination or internal initiation (Thorne and Goodfellow 2014). The RdRp functions in combination with cellular components and comprises of palm, finger and thumb domains (Shatskaya and Dmitrieva 2013). Overall this forms a right hand structure in which two catalytic aspartate AAs are contained in the conserved palm domain, whilst the thumb and fingers surround the template to produce a channel for replication (Ortin and Parra 2006).

1.5.2 Virion structure

A NoV genome is encased within a virion by 180 major capsid molecules (VP1) (Michael G. Rossmann 1989; Prasad et al. 1994b). The VP1 molecules interact to form 90 dimers, and from the centre can be separated into a contiguous shell (S domain) and an arch-like protruding domain (P domain) (Prasad et al. 1999; Prasad et al. 1994b). The P domain dimers are present at all local and strict two fold axes, generating a hollow at each icosahedral five- and threefold axis (Prasad et al. 1999; Prasad et al. 1994a; Prasad et al. 1994b) (Figure 1.7).

In the NV capsid, a single VP1 molecule is 531 AA residues in length (Jiang et al. 1992b). From the N-terminal to the C-terminal the S domain consists of AAs 50-225, whilst the latter 226-520 residues form the P domain. Each VP1 molecule is identical but arranged into 3 quasi-equivalent units (A, B and C) based upon their

environment. Multiple VP1 dimers alternate between “flat” and “bent” conformations to generate an icosahedral shell (C/C and A/B, respectively) (Figure 1.7) (Prasad et al. 1999). In NV, at the N terminus, two of the three quasi-equivalent units are disordered (Prasad et al. 1999). The N termini, in an unfixed state, have been proposed to act as a switch between conformations, but are not essential for virus-like particle (VLP) formation (Bertolotti-ciarlet et al. 2002; Prasad et al. 1999). A hinge region located between the S domain and P domain imparts flexibility and could allow these interactions to influence the surface structure (Chen et al. 2006; Prasad et al. 1999).

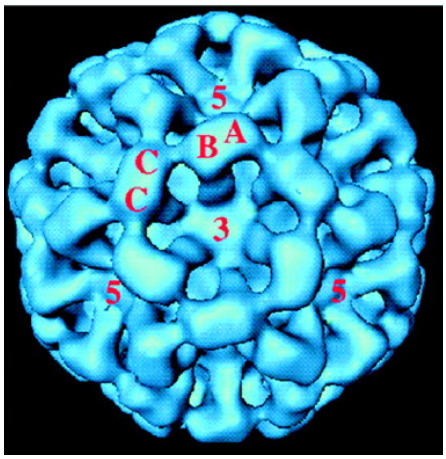


Figure 1.7 The three-dimensional structure of NV, the 3 quasi-equivalent units and axes of symmetry are labelled. Used with permission from V.Prasad (Prasad et al. 1999).

The S domain is the most conserved capsid component, and folds into an eight-stranded anti-parallel β sandwich (Bertolotti-ciarlet et al. 2002; Prasad et al. 1999). The primary function of the S domain is to form the icosahedral shell, and can independently form smooth icosahedral particles, in a baculovirus expression system, if the VP1 molecule is truncated to remove the P domain (Bertolotti-ciarlet et al. 2002).

The P domain can be split further into the P1 (arch leg) and P2 (globular head) subdomains (Chen et al. 2004; Prasad et al. 1999). In the prototype NV, VP1 amino acids 279-405 are an insertion, the P2 subdomain, surrounded by residues 226-278 & 406-520, the P1 subdomain (Prasad et al. 1999). X-ray crystallography of NV (GI.I) and VA387 (GII.4) NoVs has shown the P2 subdomain consists of 6 β antiparallel sheets (Cao et al. 2007; Prasad et al. 1999). Whereas the P1 subdomain contains a single α helix, a small β sheet and a larger β sheet, forming a stem, with a hydrophobic core and a globular head (Cao et al. 2007; Prasad et al. 1999). Interactions between the P1 subdomain and S domain are determinants of particle size, whilst dimeric interactions of the P domain increase capsid stability (Bertolotti-ciarlet et al. 2002). The P2 subdomain is responsible for interactions with the host cell receptors and NoV antigenicity (Bertolotti-ciarlet et al. 2002; Cao et al. 2007; Tan et al. 2003).

A basic minor structural protein (VP2) is also present in the prototype NV, although VLPs can be formed in the absence of this structural protein, its co-expression generally improves the yield and stability of recombinant VLPs (Bertolotti-ciarlet et al. 2002; Glass et al. 2000). VP2 is hypothesised to be positioned inside the capsid and to have roles in regulating assembly during replication, genome binding and increased particle stability (Bertolotti-Ciarlet et al. 2003; Glass et al. 2003; Glass et al. 2000). The number of VP2 molecules per NV VLP has been estimated to be ~1.5 (Glass et al. 2000). The size of the VP2 protein associated with VLPs is the expected protein size (23 kDa), but in whole virus in stool it is larger (35 kDa) and implies that

the use of an insect cell expression system or the absence of the NoV genome influences VP2 gene regulation (Bertolotti-Ciarlet et al. 2003).

1.5.2.1 Structure and function of the P domain

In the P domain, groups of AAs act in concert to form epitopes on the virus surface. Numerous residues in the P1 and P2 sub domains have been proposed to influence the surface structure, and most have been mapped onto GII.4 NoV to understand the role of herd immunity as a driver of antigenic diversity, emergence and spread of variant strains (de Graaf et al. 2016) (Figure 1.8). The P2 region is the most exposed region on the surface of the conformational NoV capsid, and thereby most likely to be involved in direct interactions between host receptors and Abs (Cao et al. 2007; Prasad et al. 1994a).

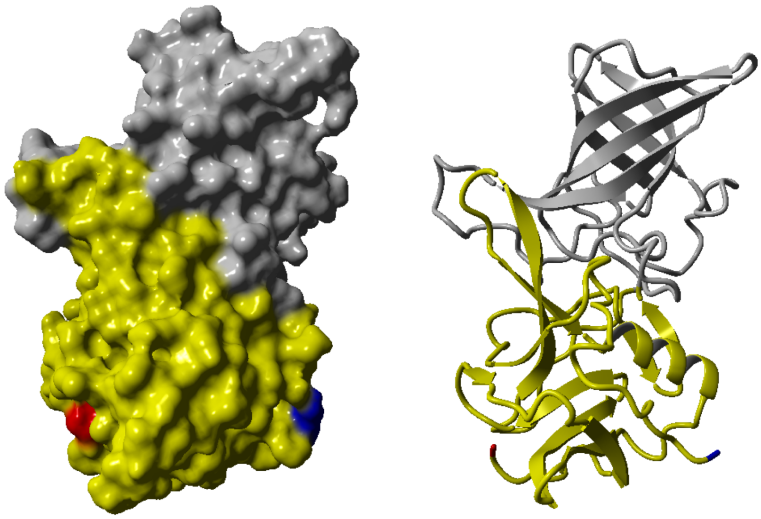


Figure 1.8 Norovirus P domain of a GII.4 consensus sequence; the colours represent protein subdomains and termini (Blue=N-terminus, Red=C-terminus, Yellow=P1 subdomain, Grey=P2 subdomain) mapped to X-ray crystallography data from the Protein Data Bank (PDB): 4OOS (Singh et al. 2015).

HuNoV VP1 sequence data is recorded routinely by different surveillance systems, and has been used to identify conserved protein changes that could define a phenotypic change. Chakravarty and colleagues utilised evolutionary trace analysis which, unlike conventional phylogenetic methods, partitions nodes on a tree by conserved class specific residues (Chakravarty et al. 2005). This meant conserved protein sequence differences could be identified between genus members that could influence host receptor binding patterns or epitope structure (Chakravarty et al. 2005). In HuNoV GII.4, 15 years of capsid surveillance data enabled the identification of residue positions which are hotspots of variation, of which significantly more (24% of total) occurred in the P2 domain (Siebenga et al. 2007).

On the other hand, sites of variability in the HuNoV structural genes have been examined on an individual level, mainly in immunocompromised patients as a putative reservoir of epidemic variants. Several of these studies have reported a higher proportion of non-synonymous mutations in the P2 domain, however whether these changes affect epitopes would depend upon the presence of selective pressure (Carlsson et al. 2009; Kundu et al. 2013; Nilsson et al. 2003). In a separate investigation, the HuNoV VP1 gene was predicted to have less positively selected sites than VP2 in the immunocompromised host (Vega et al. 2014b). However, in the same investigation, the P2 subdomain had significantly more mutations, and 3 of the 5 sites with the greatest entropy corresponded to AAs which form viral epitopes recognised by specific antibodies (Vega et al. 2014b).

1.5.2.2 Monoclonal antibody and phage display mapped sites

Several combinations of AAs have been proposed as epitopes of GII.4 HuNoV, and elucidated either *in silico* or inferred from seroepidemiological studies. The biggest limitation for confirming such epitopes remains the lack of a reliable cell culture or animal model system. Two sites in the P2 subdomain, consisting of 3 AAs each 296-298 and 393-395 (Epitope A and D, Table 1.2), were identified to mutate alongside the emergence of GII.4 epidemics (Allen et al. 2008). Furthermore, separate murine monoclonal antibodies (mAbs), isolated after inoculation with parental GII.4 VLPs from different epochs, targeted their respective motifs at epitopes A and D in chimeric VLPs, implying both were immunogens (Allen et al. 2009). Another study described Epitope D as having an additional role in HBGA binding specificity by measuring the affinity of time-ordered VLPs to carbohydrates (Lindesmith et al. 2008). The role of epitope D in antigenicity was further emphasized by investigations which examined the ability of mAbs, recovered from mice immunised with GII.4-1987 and GII.4-2006 VLPs, to neutralise chimeric VLPs (Debbink et al. 2012). In the same study, assays measuring VLP carbohydrate binding affinity, demonstrated the influence epitope A has upon HBGA binding patterns (Debbink et al. 2012).

Subsequently, five groups of residues (epitope A, B, C, D and E) were proposed as GII.4 NoV epitopes, however only two (A and D) were antigenically mapped by human mAbs (Lindesmith et al. 2012a) (Table 1.2 and Figure 1.9). Epitope E was demonstrated to be a strain-specific (GII.4-2002) blockade epitope by a human mAb (Lindesmith et al. 2012b). Most epitopes are situated near conserved receptor-

binding sites or adjacent to residues that determine specificity (Figure 1.9 and 1.10), and immune driven selection of epitopes has been proposed to generate novel GII.4 strains, that escape herd immunity, and have altered binding patterns to host receptors (Cao et al. 2007; Lindesmith et al. 2008; Tan et al. 2008).

Isolation of a human mAb capable of neutralising several GII.4 strains has meant a conserved epitope across the genotype could exist, however experiments to identify the epitope revealed a group of residues (NERK motif – Epitope N) that could regulate conformational access to the epitope allosterically (Lindesmith et al. 2014). More recently, phage display was used to demonstrate an interaction between a mAb and positions 397 and 447, whereas mutation of an adjacent site, 396, altered VLP HBGA recognition pattern (Carmona-Vicente et al. 2016).

Table 1.2 Mapped epitopes on the GII.4 P domain in the literature (Each colour corresponds to the epitope position on the P domain model below).

	VP1 amino acid position																																											
Publication	226	294	296	297	298	310	316	331	333	336	338	340	343	344	345	346	347	348	368	372	373	374	376	382	389	390	391	392	393	394	395	396	397	407	412	413	440	441	442	443	444	448	484	493
Epitope		A				N			B			C							A				C	B			D		D					E									N	
Allen et al, 2008																																												
Debbink et al, 2011																																												
Lindesmith et al, 2012																																												
Lindesmith et al, 2014																																												
Vicente, et al 2016																																												

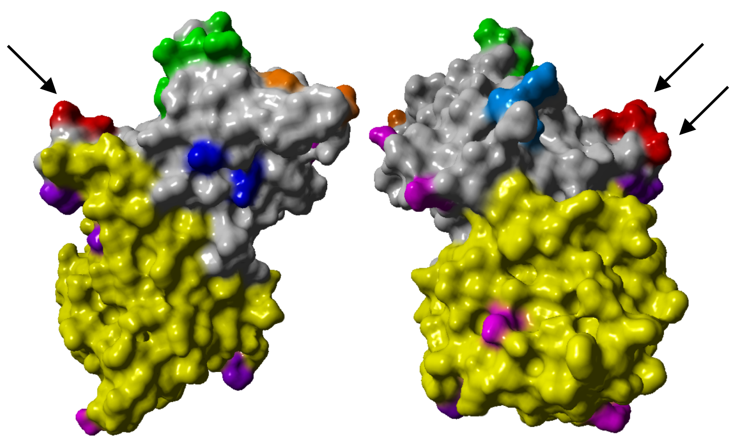


Figure 1.9 Epitope residue mapping of a GII.4 P domain mapped to X-ray crystallography data from the Protein Data Bank (PDB): 4OOS (Singh et al. 2015). The arrows show sites mapped as HBGA interaction sites and epitopes.

Table 1.3 Mapped HBGA interaction sites on the GII.4 P domain in the literature (Each colour corresponds to the position of HBGA interaction sites on the P domain model below).

Publication	VP1 amino acid position																				
	226	294	296	297	298	310	316	331	333	336	338	340	343	344	345	346	347	348	368	372	373
Cao et al, 2007														red	red	red		red			
Tan et al, 2008								green		blue	blue		blue				blue	green			
Vicente et al, 2016																					

Publication	VP1 amino acid position																						
	374	376	382	389	390	391	392	393	394	395	396	397	407	412	413	440	441	442	443	444	448	484	493
Cao et al, 2007	red				pink	red	pink	pink		pink						red	red	red	red	red			
Tan et al, 2008				green																			
Vicente et al, 2016											cyan												

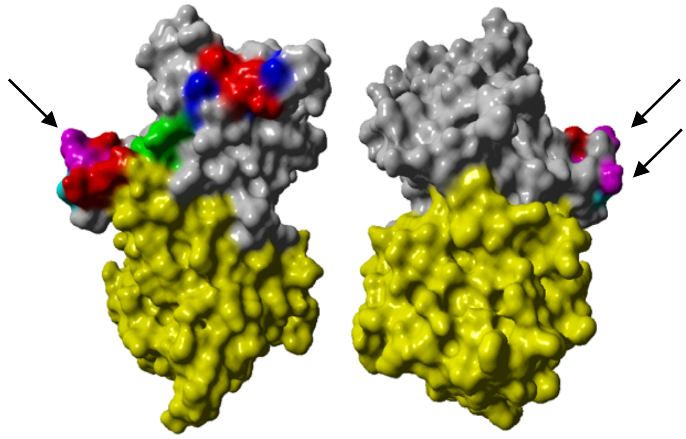


Figure 1.10 HBGA residue mapping of a GII.4 P domain mapped to X-ray crystallography data from the Protein Data Bank (PDB): 4OOS (Singh et al. 2015) - (Red=Integral to HBGA binding, Blue=Influences binding specificity, Green=Supports binding site, Pink=Second binding cavity). The arrows show sites mapped as HBGA interaction sites and epitopes.

1.6 Virus replication

The lack of a reliable culture system has hampered our understanding of the HuNoV replication cycle. Most of the information available has been derived from analogous systems, using virus relatives (e.g FCV or MNV). It is noteworthy that MNV infects macrophages and are principally associated with pathology of the neurological and immune systems, rather than enteric disease (Karst et al. 2003; Wobus et al. 2004).

1.6.1 Attachment and entry

Attachment to the host cell receptor is the first stage of a virus replication cycle, followed by internalisation. NV binding and internalisation was demonstrated first in the Caco-2 cell line (Colon adenocarcinoma tissue) (White et al. 1996), and led to

the identification of H type 1 or type 3 HBGA expression as a factor in virus binding to Caco-2 cells or duodenal tissue section (Marionneau et al. 2002). A variety of HBGA binding patterns have been shown experimentally for different HuNoV genotypes, and has provided further evidence for their role as receptors or co-receptors (Huang et al. 2005). Although HuNoV-HBGA interactions are well defined in the literature, virus cell attachment can occur in their absence and this is evidenced by heparan sulphate binding (Murakami et al. 2013; Tamura et al. 2003).

Functional receptors and mechanisms of entry are more comprehensively understood in FCV and MNV, since tissue culture models were established much earlier than HuNoV (Kreutz et al. 1994; Wobus et al. 2004). FCV was described to bind N-linked glycans on the cell surface, whilst Junctional Adhesion Molecule 1 permits cell entry and clathrin-mediated endocytosis, the virus genome is then released into the cytoplasm by endosome acidification (Kreutz et al. 1994; Makino et al. 2006; Stuart and Brown 2006). MNV, strain CR3, shares an affinity for N-linked glycans with FCV whilst other strains in this genogroup attach to gangliosides (Taube et al. 2012; Taube et al. 2009). After attachment, entry can occur via two recently elucidated functional receptors belonging to the CD300 family (Haga et al. 2016; Orchard et al. 2016). In contrast to FCV, MNV entry into the cytoplasm is pH independent, but reliant upon cholesterol and dynamin II (Gerondopoulos et al. 2010; Perry and Wobus 2010).

1.6.2 Replication

Once a NoV genome is released into the cytoplasm, RNA-protein interactions occur between the 5’ and 3’ secondary structures of the viral genome and host cell

factors. Putative interactions with PTB, DDX3 and La have been reported to have roles in viral translation and replication (Vashist, Urena et al. 2012). To begin translation the 5' covalently bonded Vpg protein interacts with the eIF4E translation initiation factor to assemble the ribosomal preinitiation complexes for translation (Daughenbaugh et al. 2003; Goodfellow et al. 2005; Herbert et al. 1997) (Figure 1.11).

The ORF1 produces a polyprotein that is cleaved into non-structural proteins of the replicase complex (Donaldson, Lindesmith et al. 2008). Viral 3C-like protease is responsible for auto cleavage of this polyprotein into the RNA-dependent RNA polymerase, $3CL^{pro}$, Vpg, p22, NTPase and p48 (Hardy 2005). More specifically, mapped ORF1 cleavage sites include Q^{330}/G^{331} and Q^{696}/G^{697} corresponding to the N terminal region, NTPase and the p20VPgProPol precursor which is hydrolysed into its components at E^{875}/G^{876}, E^{1008}/A^{1009} and E^{1189}/G^{1190} (Belliot, Sosnovtsev et al. 2003) (Figure 1.11).

The lack of tissue culture systems for HuNoV, prior to the recently described B cell and enteroid based systems (Ettayebi et al. 2016; Jones et al. 2014), means that most investigations into virus replication have been performed in feline calicivirus (FCV) and murine norovirus (MNV) model systems. In the FCV model system, virus infection induced internal membrane rearrangements and vesicle accumulation in Candell Reese Feline Kidney (CRFK) cells (Green et al. 2002). Further investigations localised enzymatically active replication complexes to the membranous pellet and an increase in the levels of negative sense templates of FCV genomic and subgenomic RNA (Green et al. 2002). Internal membrane rearrangements have also

been observed, by EM, during MNV infection of RAW 264.7 macrophage cells, which led onto organelle displacement and a loss of the golgi apparatus (Wobus et al. 2004). Immunofluorescent microscopy demonstrated the vesicles formed by MNV infection contained early endosome, golgi body and endoplasmic reticulum markers, which indicated these organelles were sites of early cell membrane recruitment by the virus (Hyde et al. 2009). Further investigations, with fluorescently labelled β-tubulin, vimentin, the MNV polymerase and dsRNA intermediates, were able to associate virus induced vesicles with the microtubule organising centre; and the role of microtubules in their subsequent accumulation in a region adjacent to the nucleus (Hyde et al. 2012; Hyde et al. 2009). The association of this positive-sense RNA virus and its replication complexes with vesicle structures is in agreement with previous studies, and is thought to have roles in protection, confinement and stability (Mackenzie 2005).

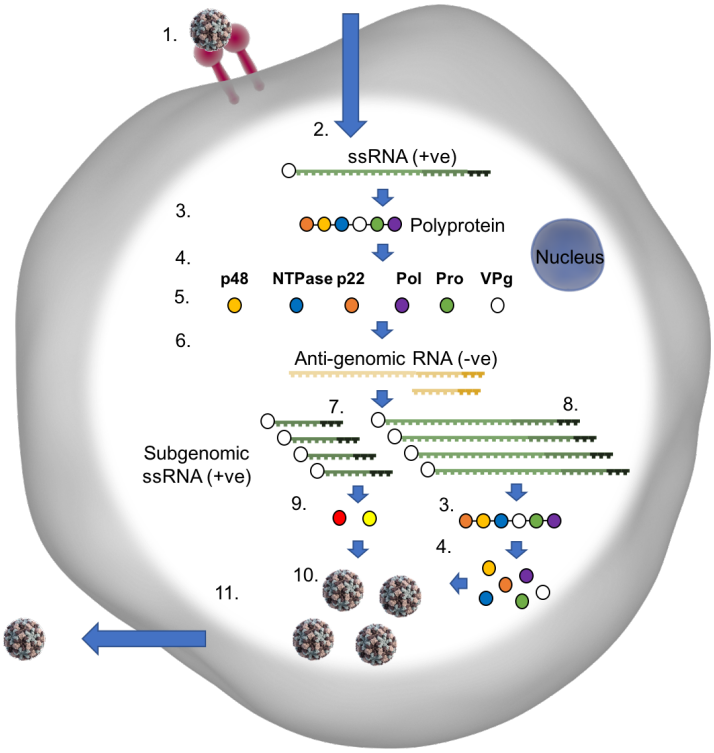


Figure 1.11 *Caliciviridae* replication adapted from Antiviral Research (Rohayem, Bergmann et al. 2010). Virus attachment (1), uncoating and entry (2), (+) ssRNA translation into a polyprotein (3), protease cleavage of the polyprotein into non-structural proteins (4), assembly of the non-structural proteins into the replication complex (5), formation of the anti-genome from the genomic RNA (6), the anti-genome is utilised as a template for both genomic and subgenomic RNA synthesis (7, 8), the subgenomic RNA is then translated into structural proteins (9), non-structural proteins are released from the polyprotein and packaged alongside subgenomic structural proteins and genomic RNA. Genomic RNA and structural proteins are assembled into mature virions (10) which exit and infect other cells, ultimately to be transmitted (11).

1.7 Host receptors and susceptibility

HuNoVs can bind to host histo-blood group antigens (HBGAs), and these molecules are associated with susceptibility to infection and disease. These carbohydrate

molecules are present on the surface of red blood cells and the mucosal epithelium (Ravn and Dabelsteen 2000). As well as this, HBGAs are found as free oligosaccharides in saliva, blood and milk of other secretor individuals (Ravn and Dabelsteen 2000). In the small intestine, these molecules are thought to act as receptors or co-receptors for HuNoV infection. HBGAs are genetically variable, and thus HuNoV binding affinity is relative to type (Table 1.4). These molecules are glycoproteins formed from a type 1 or type 2 precursor, and the sequential addition of a fucose or terminal galactose/N-acetylgalactosamine by fucosyl transferase enzymes (Figure 1.12) (Henry et al. 1995).

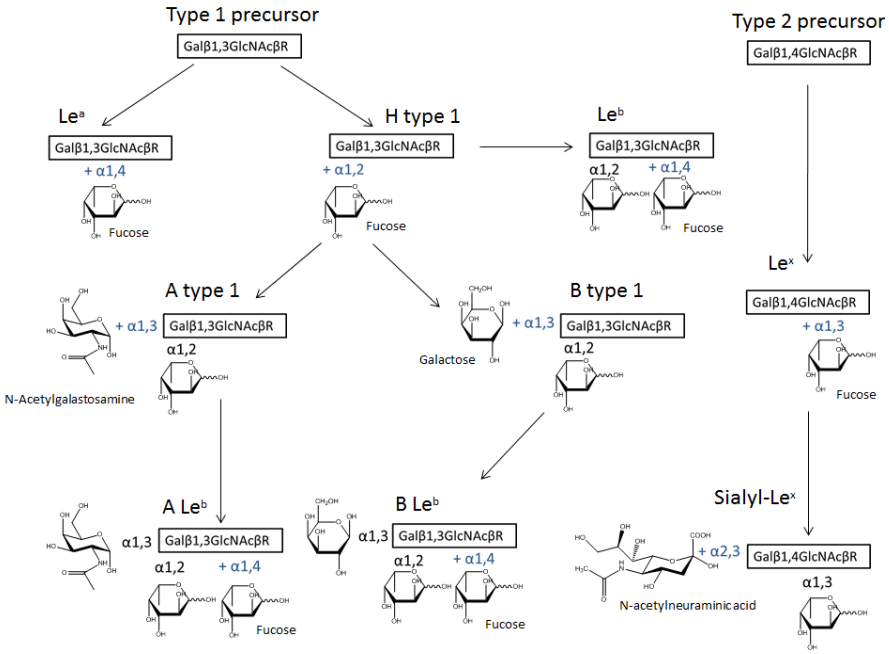


Figure 1.12 Type 1 and type 2 human histo-blood group antigens adapted from Reviews in medical microbiology (Ruvoen-Clouet, Belliot et al. 2013).

Individuals with a nonsense mutation within the FUT2 gene, which encodes the enzyme responsible for synthesising these molecules, are resistant to HuNoV infection due to an absence of the receptor (non-secretors) (Marionneau et al.

2002). AAs in the HBGA binding sites are highly conserved, though mapped proximal areas are thought to influence antigenicity implying receptor affinity and immune escape could be related (Allen et al. 2009; Debbink et al. 2014; de Rougemont et al. 2011).

HuNoV resistant patient subsets were identified much earlier than the relevance of HBGA molecules as potential norovirus receptors (Parrino et al. 1977). In this study, Parrino and colleagues identified three volunteers that were found to be resistant to Norwalk agent, with a decreased immune response relative to other individuals, implying a role for genetics (Parrino, Schreiber et al. 1977). Through assaying the HBGA molecules in saliva against blood group it is possible to identify structural binding patterns between genogroups and host phenotypes (Huang, Farkas et al. 2003) (Table 1.4).

Susceptibility to viruses is dependent upon tissue tropism and whether host cell entry can occur via binding of surface receptors. A receptor type for the *Caliciviridae* was first identified by applying Rabbit haemorrhagic disease virus (RHDV) VLPs to human erythrocytes and rabbit tissues to deduce the importance of HBGA surface molecules (Ruvoen-Clouet, Ganiere et al. 2000). From this basis it was possible to prove that H type HBGAs on the surface of Caco-2 cells contribute to virus internalisation in secretor positive individuals, and are not decoy receptors (Marionneau, Ruvoen et al. 2002).

Table 1.4 HBGA binding patterns of different HuNoV genotypes adapted from Journal of virology (Huang et al. 2005).

Genogroup / Genotype		HBGA binding pattern		
		A/B	H	Le
I	1	+++	++	-
	2	+++	++	-
	3	-	-	-
	8	-	++	++
II	1	++	+	-
	2	+++	-	-
	3	+++	+	-
	4	+++	++	-
	5	+++	-	-
	9	-	+	++
	21	-	-	++

1.8 Pathogenesis

Symptoms present during clinical norovirus infection include acute onset of vomiting, diarrhoea, nausea, abdominal cramps and myalgia (Patel, Hall et al. 2009). To increase understanding of how the virus induces these host responses histological, biochemical and physical studies have been required in both human volunteers and animal models (Karst 2010).

Early histological studies identified altered morphology of jejunal tissues during infection including shortening and broadening of villi, disorganisation of cells lining the epithelium, the presence of mucosal lesions, inflammatory cell infiltrate and widening of intercellular spaces due to oedema (Agus et al. 1973; Dolin et al. 1975). Both asymptomatic and clinically infected volunteers show malabsorption of D-xylose, lactose and lipids upon administration of virus inoculum (Blacklow, Chanock et al. 1972). However, these results could be influenced by intestinal morphology or variations within individual microbiomes.

Immune cell infiltrates have been identified to be CD8+ T cells expressing perforin leading to cell induced apoptosis of enterocytes (Karst 2010). Despite a significant increase of cell proliferation biomarkers (Ki-67) in epithelial cell tissue, HuNoV infection alters gut physiology and could be due to virus mediated apoptosis, cytokines produced by CD4+ T cells, a reduction in tight junction expression increasing permeability and a cyclic adenosine monophosphate (cAMP) independent anion secretion mechanism (Troeger, Loddenkemper et al. 2009).

Various animal models exist to better understand the NoV disease process in humans, including mice, humanised mice, Gnc pigs, calves and chimpanzees (Karst, Wobus et al. 2003, Cheetham, Souza et al. 2006, Bok, Parra et al. 2011). To obtain an accurate reflection of pathogenesis the animal model used must be like the host of interest. This also applies to the virus used, hence why the discovery of MNV-1 was an important development. The discovery of MNV-1 allowed for the first norovirus tissue culturing system, and can be applied to transgenic mice models (Wobus, Karst et al. 2004, Wobus, Thackray et al. 2006). Mice do not have an emetic reflex, instead gastric bloating appears and could correlate with gastric infection and emesis in humans (Kahan, Liu et al. 2011). An injury to the digestive tract is thought to be induced by delayed gastric emptying, leading to vomiting in a proportion of individuals (Meeroff, Schreiber et al. 1980). Other models can remove microbial influence by using Gnc animals born in aseptic conditions with controlled exposure to microorganisms. Infected Gnc pigs and calves are proven to shed GII.4 viruses (Cheetham, Souza et al. 2006). Gnc calves infected with GII.4 virus show less

severe pathologies when compared to bovine norovirus infection (Souza, Azevedo et al. 2008).

Problems with these models can include the induction of morbidities not present within human clinical infection. Specifically, MNV-1 symptomatic infection can only occur in STAT1$^{-/-}$ mice, whereas infection of wild type mice is limited to a mild inflammatory response (Mumphrey, Changotra et al. 2007). After infection of transgenic mice, the virus utilises dendritic cells (DC) to transport itself to distal sites such as the mesenteric lymph nodes (Elftman, Gonzalez-Hernandez et al. 2013). Though initial infection takes place in the intestine, the latter stages are non-localised, whereas HuNoVs usually remain in the gut.

1.9 Detection and diagnostics

Prior to the widespread availability of molecular or serological diagnostic tools (Table 1.5), the Kaplan criteria were developed to distinguish HuNoV infection from other aetiological agents of GE. The Kaplan criteria discerns HuNoV outbreaks if i) the mean duration of illness is 12-60 hours ii) the mean incubation period is 24-48 hours iii) vomiting is present in > 50% of individuals and iv) bacterial pathogens are absent in stool (Kaplan et al. 1982). Whilst this was an effective and reasonably sensitive tool for the identification of HuNoV outbreaks, it was not able to identify sporadic cases. Currently various molecular and serological assays are available for the laboratory diagnostic of HuNoV GE (Kirby and Iturriza-Gomara 2012; Vinje 2015).

During clinical manifestation of HuNoV infection, the virus can be detected in vomitus, mouthwashes, rectal swabs or stool specimens (Atmar et al. 2008; Kirby et

al. 2011; Kirby et al. 2010; Nakanishi et al. 2009). However, diagnostic tests will predominantly be performed on stool samples due to the high viral load, virus persistence in stool and the sporadic nature of vomiting during illness (Kirby et al. 2010; Rockx et al. 2002; Vinje 2015).

EM allows the detection of virus in stool samples through observation of the characteristic shape, and can therefore detect all genotypes and strains present without potential biases. However, it is a highly skilled and labour intensive method that is also relatively insensitive, and its use for diagnostic purposes is currently limited (Costantini et al. 2010; Phillips et al. 2009). EIA based assays have improved upon the sensitivity of EM by utilizing cocktails of pAbs and mAbs raised against different HuNoV VLP strains, however cross-reactivity between genogroups and the inability to detect certain genotypes has been described (de Bruin et al. 2006; Costantini et al. 2010). Numerous studies have identified cross-reactive mAbs, between and within GI and GII, which could further increase EIA sensitivity, but would decrease specificity (Hale et al. 2000; Kitamoto et al. 2002; Li et al. 2009b; Lindesmith et al. 2012b). EIA assays are limited further by an antibody losing affinity to contemporaneous strains via antigenic drift of HuNoV strains and the requirement for adequate sample storage, to prevent proteolytic degradation of epitopes (Bull et al. 2010; Costantini et al. 2010). The use of ELISA or other antigen detection systems is particularly appropriate in outbreak situations in which numerous different samples are available for testing.

The most widely used diagnostic assays for HuNoV are based upon RT-PCR (Kirby and Iturriza-Gomara 2012). Compared to EM and EIA based assays, PCR is able to

detect much lower titres of virus and therefore HuNoV infection resulting in moderate shedding can be detected and/or quantified (Amar et al. 2007; Atmar et al. 2008). Several conventional RT-PCR assays exist for the detection of NoV (Ando et al. 1995; Green et al. 1995; Vennema et al. 2002; Vinje and Koopmans 1996; Wang et al. 1994). However, these have been largely superseded by second generation qPCR techniques, and the increase in diagnostic test sensitivity has meant thresholds for asymptomatic infection or shedding of low viral loads prior to or post-illness can be better defined (Kageyama et al. 2003; Phillips et al. 2009).

Other, less established, quantitative techniques exist for HuNoV in clinical samples. LAMP (Loop-mediated isothermal amplification) and NASBA (Nucleic acid sequence-based amplification), like PCR assays, quantify HuNoV by amplicon replication, mediated by *Bst* DNA polymerase and T7 RNA polymerase, respectively (Compton 1991; Notomi et al. 2000). Amplicon replication by enzymes, other than Taq DNA polymerase, can be performed at lower isothermic temperatures, and are less susceptible to inhibition (Iturriza-Gomara et al. 2008; Lamhoujeb et al. 2009). Luminex technology, on the other hand, requires a multiplex PCR reaction prior to laser-assisted quantification of oligonucleotide coated microspheres, which can allow for high levels of multiplexing (Dunbar 2006; Liu et al. 2011). Massively parallel sequencing (MPS) also undergoes an amplification step to enrich nucleic acids prior to detection, and the read data has been shown to correlate with semi-quantitative PCR measurements (Nakamura et al. 2009).

NoV disease can also be identified with qualitative diagnostic tests, which tend to be more rapid. Gas chromatography (GC) of volatile organic compounds, in clinical

stool samples, and mass spectra of the NoV VLP spiked stool samples, have unique profiles which could help inform the diagnostic result (Colquhoun et al. 2006; Probert et al. 2004). Microarrays have also been developed to identify the presence of different genotypes in a sample (Mattison et al. 2011).

Table 1.5 Summary of current HuNoV diagnostic methods.

Diagnostic technique	Sensitivity	Advantage	Disadvantage
EM	+	Not pathogen specific	Requires a skilled technician Expensive Difficult to discern virus
IEM	++ (Limit: $\leq 10^6$ particles/g faeces) (Constantini, 2010)	Moderate increase in sensitivity and specificity	
ELISA	++ (Limit: $\leq 10^6$ gEq/g faeces) (Costantini et al. 2010)	Low technical requirements	Sensitivity not equal for all genotypes Insensitive for sporadic case diagnostics
PCR/qPCR	++++	Detection limit of one genomic copy Multiplexing	Inhibitors can generate false negatives Does not coreelate with symptoms
Metagenomics	?	Not pathogen specific	Expensive Complex data analysis Large capacity for data storage required No standard to assess the diagnostic value of result

The recent development of two complex *in vitro* cultivation methods for HuNoV have meant that a diagnostic assay to quantify infectious virus may be feasible (Ettayebi et al. 2016; Jones et al. 2014). The first to be described, can provide relatively low yields of HuNoV from BJAB cells (Human Burkitt lymphoma B cell

line), but is reliant on the presence of a cofactor in stool that is thought to be H antigen on the surface of enteric bacteria (Jones et al. 2014). In the second technique, HuNoV is replicated to greater yields in enterocytes, or intestinal crypt stem cells, and whilst this provides greater yields it also requires a high level of technical skill (Ettayebi et al. 2016). The enterocyte system reflected more closely the virus biology, as GII.3 was shown to infect the non-secretor phenotype whereas GII.4 could not, but also bile was found to be either a required cofactor or enhancer of virus replication for which the mechanism is not well understood (Ettayebi et al. 2016).

1.10 Norovirus immunity

In the first NV and NLV human challenge studies, short term immunity, to a homologous strain, was present 9-14 weeks' post exposure (Wyatt et al. 1974). In subsequent experiments, a subset of individuals were not infected after repeat inoculations, whilst the remaining individuals did not appear to be immune 2-3 years after the initial challenge (Parrino et al. 1977). However, the inoculum used for these challenge studies was found to far exceed either of two proposed ID_{50} values for NV (Atmar et al. 2014; Teunis et al. 2008).

Correlates of innate immunity are not well understood in HuNoV infection. IFN expression has been found to coincide with oral VLP inoculation of humans (IFN-γ) (Tacket et al. 2003), and decreased levels of NV RNA in a cell line replicon system (IFN-α) (Chang et al. 2006). Furthermore, in the MNV model system, both DCs (Elftman et al. 2013) and the signal transducer and activator of transcription 1

(STAT1) pathway, influenced by IFN or extraneous IFN-λ treatment, were shown to control or abolish virus infection *in vivo* (Mumphrey et al. 2007; Nice et al. 2015).

An infected individual was found to be symptomatic over a 2 year period with a low cellular immunity and an intact humoral immunity (Nilsson, Hedlund et al. 2003). This implies cellular immunity is necessary and contributes to complete viral clearance. Another example of the broad immune response required was present in a volunteer based study that could segregate genetically susceptible individuals with varied norovirus IgA specific responses relative to total IgA (Lindesmith et al. 2003). In this study two subsets of resistant individuals were identified; those with an early IgA response above base line during the first five days after infection (infection was defined as the detection of viral RNA in stool by RT-PCR or ≥ 4 fold increase in Norwalk virus-specific serum), and those resistant with a lesser or no secretory antibody response, which implied resistance is not solely dependent upon humoral immunity (Lindesmith, Moe et al. 2003).

1.10.1 Correlates of human norovirus immunity

In the literature, further HuNoV challenge studies and vaccine development trials have been performed to identify immune correlates that constitute protection from virus infection and disease (Ramani et al. 2016).

1.10.1.1 Host genetics

In 1977, Parrino and colleagues proposed the presence of a genetic factor that could influence whether individuals were infected by NV (Parrino et al. 1977). The risk of infection was identified to be related to the ABO group system (Hutson et al.

2002), and consequently investigations found an inactive α(1,2)fucosyltransferase (FUT2) to be an indicator of NV resistance (Lindesmith et al. 2003). The FUT2 enzyme, as described previously, is involved in the synthesis of HBGAs on the surface of epithelial cells (Henry et al. 1995), however, whether HBGAs function as a receptor or co-receptor is not known.

1.10.1.2 The adaptive immune system

Most of the recognised factors that correlate with protection from HuNoV are a part of the adaptive immune system, and this likely signifies its importance in virus clearance (Ramani et al. 2016). In 2003, Lindesmith and colleagues found a more rapid salivary IgA response (response within the first five days in contrast to a peak titre on day 14), after NV challenge to indicate protection in susceptible individuals (Lindesmith et al. 2003). This agrees with research from a recent challenge study, which found the presence of pre-existing NV specific salivary IgA, alongside specific IgG memory B cells, to be associated with protection from virus induced GE (Ramani et al. 2015). In the same study, individuals with larger NV-specific IgA titres in stool tended to shed a lower peak viral load, and on day seven (post-challenge) this was found to correlate with a decrease in shedding duration, but it was the presence of NV-specific IgG memory cells prior to challenge that correlated with protection (Ramani et al. 2015). Several studies have noted a relationship between the ability of patient sera, prior to exposure, to block HuNoV HBGA interactions or virus induced red blood cell (RBC) haemagglutination to correlate with an absence of infection and illness (Atmar et al. 2015; Atmar et al. 2011; Czako et al. 2012; Reeck et al. 2010). Moreover, in one of these studies GII.4 specific IgA in sera was shown

to be an indicator of resistance to infection and illness, in individuals receiving placebo rather than a vaccine (Atmar et al. 2015).

1.11 Treatment and prevention

Several strategies exist to nullify HuNoV infection and transmission; including infection control, vaccination, antivirals and immunotherapy. Infection control constitutes preventing and limiting outbreaks in semi-closed environments through policies on organisational structure, food-handling, hygiene, cleaning and, in healthcare settings, patient isolation and cohorting (Barclay et al. 2014). The development of an efficacious vaccine has the potential to prevent illness and reduce secondary transmission in groups at high risk of exposure (Bartsch et al. 2016). On the other hand, immunotherapy and antivirals are a necessary alternative for the treatment of immunocompromised individuals that would be unable to clear infection otherwise, or can provide a prophylactic option (Gairard-Dory et al. 2014; Rocha-Pereira et al. 2015; Woodward et al. 2015).

1.11.1 Infection control

A HuNoV outbreak is triggered by the virus being introduced into a susceptible population, in a semi-enclosed environment, by a person or vehicle; therefore food-handling and hygiene are important in prevention. To minimise the occurrence of virus introduction, individuals in healthcare environments should wash their hands with soap and warm water for at least 20 seconds, and food-handlers should avoid bare hand contact with food (Blaney et al. 2011; Division of Viral Diseases et al. 2011; US FDA 2013). Furthermore, HuNoV infected healthcare workers or food-

handlers, should be absent from work until 48 hours after cessation of diarrhoea and/or vomiting (Hall et al. 2012; Thornley et al. 2013).

If an outbreak occurs, an overarching organisational structure is required to implement measures to identify and remove the source of contamination and decontaminate the environment. In foodborne outbreaks the source could be a food-handler or commodity, whereas in a healthcare environment it is most likely a patient or healthcare worker. In nosocomial outbreaks, secondary transmission is further contained by disposable personal protective equipment (PPE), isolation of infected cohorts, designation of nurses to cohorts and limiting visits or patient transfers (Haill et al. 2012; Illingworth et al. 2011; MacCannell et al. 2011). Once an outbreak has been contained, it is necessary to disinfect surfaces, for which freshly prepared 1000 parts-per million (ppm) of sodium hypochlorite is recommended, and has been demonstrated to be more effective in comparison to alternative options (Tung et al. 2013).

1.11.2 Vaccination

At present, there is no vaccine available for HuNoV disease prevention. However, given the burden of disease the development of an effective vaccine is highly desirable. In the US population, if development of a HuNoV vaccine costed $25 (50 % minimum efficacy) Markov modelling predicted cost-effectiveness in groups with the greatest disease severity (> 65 years) or burden (< 5 years), if immunity lasted at least one year (Bartsch et al. 2012). However, an efficacious vaccine would need to provide protection against the continual divergence of GI and, more so, GII viruses of which estimates on the longevity of protection from vaccination or challenge are

highly variable (Cubitt et al. 1998). Hence, a bivalent HuNoV intramuscular vaccine has been developed, which consists of NV VLP and a consensus GII.4 VLP and can elicit broad immune responses to different GII.4 strains (Parra et al. 2012).

The duration of homologous immunity has been estimated to be ~4 – 8 years by mathematical modelling, and therefore could be a viable intervention method (Simmons et al. 2013). This observation markedly differs from experimental evidence, which have described the duration of immunity to be greater than six months but less than two years (Johnson et al. 1990; Parrino et al. 1977), which is further supported by epidemiological studies in infants (Blazevic et al. 2016).

In total, two delivery routes have been trialled in HuNoV immunogenicity clinical trials; VLPs have been administered intranasally or intramuscularly to cohorts aged between 18 and 50 with superior results obtained with the intramuscular administration (Atmar et al. 2011; Bernstein et al. 2015).

In a challenge clinical trial, Atmar and colleagues could reduce disease severity by intranasal immunisation with a homologous VLP prior to NV challenge (Atmar et al. 2011). A subsequent study significantly reduced disease severity, but not infection rate, in volunteers immunised intramuscularly with a mixture of NV and a GII.4 consensus VLP (Bernstein et al. 2015). The adjuvants present in the most recently developed HuNoV vaccine, aluminium hydroxide and monophosphoryl lipid A (MPL), could increase antibody titre and the number of B cells in the human papilloma virus vaccine (HPV) (Schwarz and Leo 2008). However, the data provided by these trials is limited, as vaccine induced immune responses have not yet been examined in the age groups that would benefit most (infants and the elderly).

1.11.3 Antivirals

An antiviral can provide a prophylactic or immediate treatment option for individuals that have not been vaccinated or those that are not capable of mounting an adequate immune response to clear HuNoV. The mechanism of an antiviral can target a host factor, which would have a higher genetic barrier for antiviral resistance, or a virus encoded protein, which will minimise inhibition of ongoing processes in the cell. In the literature, libraries of compounds have been screened *in silico* or *in vitro* for affinity to a NoV protease or polymerase (Eltahla et al. 2014).

Suramin and, the analogous compound, NF023 were found to have an affinity for the MNV and HuNoV RdRps by *in silico* analysis, and thus inhibited virus replication *in vitro* (Mastrangelo et al. 2012). Fragments of these molecules were used to identify a *novel* target site in the RdRp thumb domain (Tarantino et al. 2014). This led to the development of PPNDs (pyridoxal-5'-phosphate-6-(2'- naphthylazo-6'-nitro-4',8'-disulfonate) tetrasodium salt), a molecule which inhibits the RdRp by double stacking in the thumb domain (Croci et al. 2014). Instead of screening for novel inhibitors, existing antivirals for other RNA viruses can be tested for efficacy against NoVs. Favipiravir and Ribavirin, which could also be an immunomodulator, increase the rate of mutagenesis in RNA virus quasispecies to induce an error catastrophe event, of which Ribavirin has been shown to clear HuNoV chronic infection (Arias et al. 2014; Woodward et al. 2015). Another example is 2'-C-methyl-cytidine, a nucleoside analogue which has been proven to clear MNV infection, and

as a phrophylactic prevent transmission to susceptible mice (Rocha-Pereira et al. 2015; Rocha-Pereira et al. 2013).

A dipeptidyl compound (GC376), developed by Takahashi and colleagues, was shown to covalently bond to the catalytic cysteine residue within the NoV 3CL protease and form several hydrogen bonds to inhibit function (Takahashi et al. 2013). Further work on dipeptidyl compounds has identified macrocylic molecules to have differential activity against MNV, in contrast to host proteases (Damalanka et al. 2016; Galasiti Kankanamalage et al. 2015).

1.12 Context and contribution of this research

As the scientific literature over the last two decades emphasizes, HuNoVs are the most common cause of sporadic cases and outbreaks of acute gastroenteritis worldwide. Their persistence in the human population is driven by a number of factors such as virion stability, low infectious dose, a broad host binding specificity, a lack of long term immunity in the host and a high evolutionary rate (Bull and White 2011). GII.4 HuNoV is the most diverse HuNoV genotype and the most persistent; this genotype undergoes strain replacement via antigenic drift in a similar manner to Influenza A viruses.

HuNoV infection of immunocompromised individuals is widely considered to be a reservoir for the emergence of novel virus strains. In chronic HuNoV infection, virus shedding can persist for long periods, and the virus population will undergo random mutations leading to the formation of quasispecies. However, epidemiological surveillance data would suggest HuNoVs mutate at a much faster rate in the

population. Therefore, the overarching aim of this thesis is to identify the presence of quasispecies and generation of HuNoV diversity in acute infections, in which immune pressure could direct the emergence antibody escape novel strains.

1.13 The aims of chapter 2

The aim of this series of investigations was to optimise a protocol for the absolute quantification of GII HuNoV from clinical stool samples by second generation qPCR.

- To establish a reliable method for absolute quantification of HuNoV genome copy number in stool samples.
- To evaluate the impact of the use of a plasmid vs RNA standard for quantification of HuNoV in stool samples.
- To evaluate the impact of one vs two step RT-PCR and specific vs randomly primer RT-steps on quantification of HuNoV in stool samples.

1.14 The aims of chapter 3

- To evaluate the porcine-gastric mucin based HuNoV capture method as an enrichment tool for HuNoV in clinical samples, prior to MPS.
- To measure the effects of PGM-MB capture and random PCR amplification cycles on the detection and quantification of HuNoV minority variants on HiSeq (Illumina) sequence data.

1.15 The aims of chapter 4

- To identify virus populations and mutations in whole genome sequence data in clinical samples collected longitudinally from acute HuNoV infection through to convalescence, and map any changes to the protein structure.

- To accurately detect the presence of minority variants, in acute HuNoV infection, and identify whether regions of the genome are under selective pressure.
- To investigate the relationship between the presence of HuNoV specific copro-antibodies and any changes identified in the HuNoV whole genome sequence data over the course of acute infection.

1.16 The aims of chapter 5

- To validate P2 domain sequencing in HuNoV seeding events and outbreaks as a tool to discern the introduction of a single HuNoV strain or multiple virus strains in an outbreak using different sequencing technologies: Sanger and MPS.
- To identify the analytical advantages and disadvantages between a partial sequencing and whole genome sequencing approach to tracking HuNoV transmission within outbreaks.

2 Absolute quantification of GII Norovirus in clinical specimens

2.1 Introduction

2.1.1 Molecular biology and pathogen quantification

Accurate detection and monitoring virus titres is essential to various applications in clinical virology and environmental virology. Traditional methods have measured viral titres via virus-antibody (Ab) interactions, plaque forming units (pfu) *in vitro* or by high-resolution microscopy. These quantitative techniques can be limited by the availability of reagents, such as antibody's or a tissue culture model, changes in virus antigenicity over time, or the low sensitivity of microscopy and the level of experience needed to visually discern different viruses in a sample (Goldsmith and Miller 2009).

The discovery of polymerase chain reaction (PCR) and it's wide dissemination between laboratories led to the development of highly sensitive and specific molecular assays for virus quantification (Mullis and Faloona 1987; Saiki et al. 1985). The first generation of quantitative PCR (qPCR) assays differed on whether the proportion of a target was determined at the log phase of PCR or by direct competition with a separate nucleic acid at different concentrations (Competitive PCR). Quantification in the log phase of PCR compares the sample against known standards to deduce the copy number of a target, whereas the use of a directly competing nucleic acid meant target abundance could be estimated by how it's

replication is inhibited (JD 1993; Kellogg et al. 1990; Pang et al. 1990; Wang et al. 1989).

An inherent disadvantage of first generation qPCR assays was the requirement of post amplification processing, and led to the development of second generation qPCR assays (Heid Stevens, J., Livak, K.J., and Williams, P.M. 1996). The principle of second generation qPCR assays was first introduced In 1996, which utilized the 5' → 3' exonuclease activity of Taq DNA polymerase to generate fluorescence when target molecules were replicated in a closed system (Heid Stevens, J., Livak, K.J., and Williams, P.M. 1996).

Third generation qPCR assays, or digital PCR, partition a sample to amplify single genome copies in individual wells, the copy number can then be inferred by counting the number of wells with a fluorescent signal (Vogelstein and Kinzler 1999). Utilization of this technology in clinical laboratories has been limited by dynamic range and cost, however theoretically it is more sensitive than second generation assays and is less susceptible to inter-assay variability and inhibitors of amplification (Dingle et al. 2013; Sedlak et al. 2014).

The use of second generation qPCR assays is widespread amongst laboratories and has led to the need for standardization and quality controls between assays to generate meaningful results (Bustin et al. 2009; Huggett et al. 2005). To produce meaningful data that can be published, it is important to normalise the nucleic acid input, optimise amplification, monitor inefficiencies and quantify against validated standards of known concentrations (Bustin et al. 2009).

2.1.1.1 Normalization

Prior to qPCR, samples can be normalized by the sample/nucleic acid quantity or an endogenous host reference gene. To normalize by sample volume or weight the nucleic acid target must be evenly distributed in the sample. On the other hand, copy number of a target nucleic acid can be normalized to an endogenous reference gene, on the assumption it is expressed at a baseline level and is not influenced by experimental parameters (Brouwer et al. 2003).

2.1.1.2 Internal process controls

A qPCR internal process control (IC) is an endogenous or exogenous nucleic acid that can be concurrently extracted, reverse transcribed (when the target is RNA) and amplified with the target to produce a distinct signal, and thus can be examined for inefficiencies in the process.

An exogenous internal control (EIC) can be designed to compete with the qPCR target for primer binding, a competitive internal control (CIC), or undergo amplification via a separate set of primers, non-competitive internal control (NIC) (Hoorfar et al. 2004). A CIC would more closely reflect amplification of the target sequence; however, the template should be present at a lower copy number in comparison to the target and generate a longer amplicon to maintain assay sensitivity (Hoorfar et al. 2004). A NIC could compete indirectly, for enzyme and dNTPs, but this can be avoided if the reaction is not performed in a multiplex.

Simultaneously, the structure of an EIC is equally important to mimic the stability of the target at each stage of the qPCR process, and can be exposed, as nucleic acid, or encased, as armoured RNA, a recombinant virus or an animal virus surrogate

(Schrader et al. 2012). The concept of an armoured EIC was first introduced in 1998 (Pasloske et al. 1998). A portion of the HIV *gag* gene was expressed in *E. coli*, with an operator sequence to allow the self-assembly of MS2 capsid subunits around the fragment (Pasloske et al. 1998). The efficient expression of an encased RNA fragment was limited to 500 bases, but the MS2 capsid provided resistance to nuclease degradation and greater stability in plasma (Pasloske et al. 1998). A protocol to produce recombinant PR8 Influenza A virus, containing a HCV gene fragment, as an EIC was recently described and found to be more stable and cheaper than armoured RNA (Wang et al. 2016). However, live virus can be genetically unstable and infectivity would require nullification before use as an EIC, therefore closely related animal viruses tend to be evaluated instead (Gentry-Shields and Jaykus 2015).

2.1.1.3 External standards

The quantity of a qPCR target can be measured by a comparative cycle threshold (Ct) or a standard curve method. A comparative Ct method assesses the quantity of a target relative to others whilst a standard curve method measures absolute copy numbers, and hence requires an external standard (ES) of known concentration.

An ES is a nucleic acid template, consisting of identical priming sites to the qPCR target, which is quantified in parallel. In the literature, live virus, plasmids (linearized and circular), amplicons, *in vitro* transcribed RNA and synthetic deoxyribonucleic acid (DNA) standards have been validated (Bermúdez de León et al. 2013; Bowers and Dhar 2011; Dhanasekaran et al. 2010; Fu et al. 2009; Hartnell et al. 2012; Lima et al. 2017).

A validated ES should undergo reverse transcription (RT) and/or qPCR, depending upon the target of interest, and remain stable for long periods to prevent misinterpretation of sample copy number. An RNA ES will detect RT inefficiencies, and therefore sample copy number can be accurately determined retrospectively (Bowers and Dhar 2011). On the other hand, plasmid standards, and less so amplicons, have greater stability at more moderate temperatures, although there is evidence of an *in vitro* transcribed RNA ES showing no significant variability over 9 months (Bermúdez de León et al. 2013; Dhanasekaran et al. 2010). Alternatively, lenticules, or control-dried plano-convex discs, are reference materials that contain live virus and are used for quality control and quantitative purposes (Hartnell et al. 2012). Lenticules have been demonstrated to have short (1 week) and moderate (52 weeks) stability at room temperature and below -15 degrees Celsius, respectively (Hartnell et al. 2012).

2.1.2 HuNoV diagnostics

Early RT-PCR methods to detect HuNoV were mostly used qualitatively, but several semi-quantitative protocols exist for first generation qPCR of HuNoV, and some are still used, however this has largely been superseded by second generation techniques (Vennema et al. 2002). Third generation qPCR assays exist for HuNoV, but these have been used sparingly in research, and hence second generation assays are considered the “gold standard” for diagnostics (Coudray-Meunier et al. 2016; Coudray-Meunier et al. 2015; Fraisse et al. 2017; Polo et al. 2016). Although PCR based assays are the most common diagnostic tool, there are three inherent limitations to this technology. Firstly, PCR relies on a molecular process that can

generate false negative results in the presence of polymerase inhibitors; these have been comprehensively described in stool specimens and can be carried forward after the nucleic extraction stage (Schrader et al. 2012). Secondly, the high sensitivity of the RT-PCR methods means that the presence of HuNoV in a clinical sample can be detected in asymptomatic infection or for an extended period following infection and recovery due to prolonged shedding; this problem may be overcome by the establishment of an assay cut-off as a proxy for viral load when there is evidence that load is inversely related to the presence or severity of symptoms (Phillips et al. 2009). Thirdly, quantification of a genome cannot discern non-infectious HuNoV from viable virus, which can only be ascertained with a valid tissue culture system. This is still a limitation for HuNoV despite the recent description of two *in vitro* systems that sustain virus replication (Ettayebi et al. 2016; Jones et al. 2014). Whilst the ability to reproduce these methods more widely is a promising development for the HuNoV research field, at present they do not offer a realistic diagnostic tool due to their complexity and cost, in addition to the limited sensitivity to date.

2.2 Materials and methods

2.2.1 Sample preparation, RNA isolation and application of the exogenous internal control

For protocol optimization purposes, a collection of stool samples, positive for GII HuNoV, were gifted to this project from a previous study (Table 2.1). The protocol was then applied to stool samples collected from patients diagnosed with clinical HuNoV disease in the Royal Liverpool University Hospital (RLUH). Seven patients

were recruited in total, and three of these had two or more samples collected (Table 2.2).

Table 2.1 GII HuNoV archived samples used in the optimisation of an absolute quantification methodology (NA – Not applicable).

Sample ID	Genogroup	Year of detection	Ct value
A	GII	2012	21.71
B			23.36
C			35.31
N	NA		Negative

Table 2.2 GII HuNoV positive clinical samples collected for absolute quantification with the optimised protocol.

Patient ID	Genotype	Year of detection	Day collected post recruitment	GII HuNoV Ct value
D	GII.4	2017	1	24.45
			2	24.76
E	GII.16P/GII.4		0	20.03
			1	18.94
			1	19.82
F	GII.4		0	25.95
			1	18.10
			2	20.08
			3	22.54

Stool samples were suspended in sterile phosphate buffered saline (PBS) (Sigma, Dorset, UK) at 2-12% (w/w); each preparation was shaken vigorously and centrifuged at 14800 rpm for 10 minutes. An aliquot of 250 µL of the suspension supernatant was collected for nucleic acid extraction by guanidium isothiocyananate/silica method (Gn silica) (Boom et al. 1990). At the first stage of extraction, the lysis buffer was spiked with EIC template RNA (PrimerDesign, Hampshire, UK), as recommended by the manufacturer.

For lenticule validation, 2 or 3 Genogroup II Lenticule Discs (PHE-RMNOROG2/231415) were rehydrated per experiment, and this was performed three times to give seven replicates overall. A disc was rehydrated in 100 µL of sterile distilled water (SDW) for 10 minutes and vortexed thoroughly. Each disc was serially diluted from 1.1×10^{4} copies to 1.1×10^{-2} copies and the nucleic acids extracted as described above.

2.2.2 Plasmid standard curve

An *E.coli* transformant of pCRGII3-3 containing partial ORF1+ORF2+ORF3 of a GII.3 HuNoV genome (Figure 2.1) was kindly gifted by Dr D Allen (PHE). The transformant was grown overnight at 37°C in ampicillin supplemented (100 µg/mL) Luria-Bertani broth. The plasmid was extracted from the overnight culture using the Qiaprep Spin Miniprep Kit (Qiagen, West Sussex, UK), following the manufacturer's instructions. The concentration of the plasmid eluate was measured using the Qubit® DNA BR Assay Kit (Fisher Scientific, Leicestershire, UK), and then a dilution series was generated in the range of 6×10^{7} copies/µL to 6×10^{-2} copies/µL.

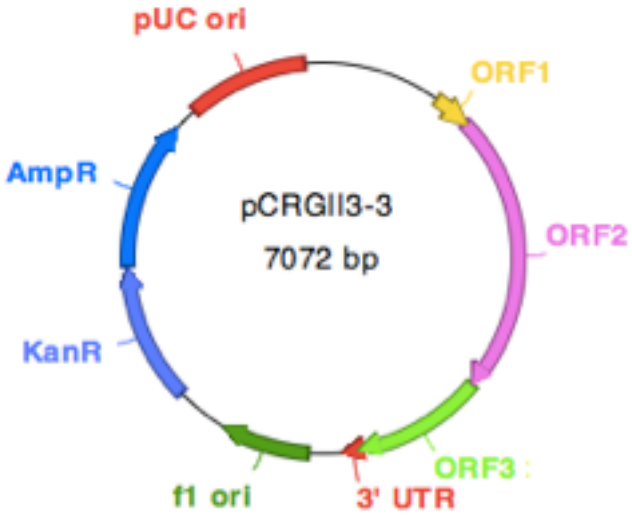


Figure 2.1 pCRGII3-3 plasmid map.

2.2.3 *In vitro* transcribed RNA curve

Primers were designed to produce an amplicon consisting of a T7 RNA polymerase promoter upstream of the ORF1/2 in pCRGII3-3 (Table 2.3). A touchdown PCR was performed using the Expand High Fidelity System (Roche, West Sussex, UK) to amplify the ORF1/2 region in 1 ng of pCRGII3-3 with 0.4 μM of the RNA SC primers (Table 2.3). The touch down PCR thermal profile was as follows: 95°C for 5 min, followed by 3 cycles of 95°C for 1 minute, 60°C for 30 seconds (decreasing by 2°C every 3 cycles, with an additional 23 cycles at 50°C), 72°C for 3 minutes and a final extension step of 72°C for 5 minutes.

The PCR product was gel-purified on a 0.9% (w/v) agarose gel (Appendix A), using the QIAQuick gel-extraction kit (Qiagen, West Sussex, UK), and pure amplicon provided a template for *in vitro* transcription with the MEGAscript T7 High Yield Transcription Kit (Fisher Scientific, Leicestershire, UK). The *in vitro* transcribed RNA

was treated with Turbo DNase (Fisher Scientific, Leicestershire UK) and precipitated using lithium chloride, as recommended by the MEGAscript T7 High Yield Transcription Kit (Fisher Scientific, Leicestershire, UK). The purified RNA was then re-suspended in 30 μL DEPC-treated water (Fisher Scientific, Leicestershire, UK). The concentration of transcribed RNA and remaining DNA template were measured using Qubit® RNA BR and DNA HS Assay Kits (Fisher Scientific, Leicstershire, UK). Subsequently, the RNA was diluted to a range of 6×10^{5} copies/μL to 6×10^{-2} copies/μL.

2.2.4 RT, One step and two step qPCR

RT reactions were performed using the SuperScript™ III Reverse Transcriptase System (Fisher Scientific, Leicstershire, UK). The RT reaction mixture (35 μL) contained 7 μL of 5 × First-Strand Buffer, 1.4 mM DTT, 5 mM $MgCl_2$, 0.29 mM dNTPs, 200 units of SuperScript™ III reverse transcriptase and 5 μL of RNA. Primer concentrations for the conversion of RNA to cDNA were as follows: for a random RT, 0.29 μM of random hexamers; for a specific RT, 0.29 μM of the COG II primers (Table 2.3); for EIC RNA, 1 μL of the Internal Control RNA RT primer mix (Primerdesign, Hampshire, UK). The thermal profile for first strand synthesis consisted of 25°C for 10 minutes, 37°C for 60 minutes and 95°C for 5 minutes.

A Rotor Gene 6000 (Qiagen, West Sussex, UK) instrument was used for one and two-step qPCR. One-step qPCR assays were performed using the 2×Precision OneStep™ qRT-PCR Mastermix (Primerdesign, Hampshire, UK). A final reaction mixture (20 μL) included 10 μL Precision OneStep™ qRT-PCR Mastermix (Primerdesign, Hampshire, UK) and 5 μL of RNA. Primer and probe concentrations

for single and multiplex reactions were as follows: for the GII assay, 0.5 μM of COG II primers and 0.125 μM of the Ring 2 P probe (Table 2.3); for the GADPH assay, 0.5 μM of primers and 0.25 μM of probe (Table 2.3); for the IC assay 1 μL of the IC primer/probe mix (Primerdesign, Hampshire, UK). The thermal profile for the one-step assay consisted of 55°C for 10 minutes, 95°C for 8 minutes followed by 40 cycles of 95°C for 15 seconds and 60°C for 60 seconds.

Two-step qPCR assays were performed using the PrecisionPLUS Mastermix (Primerdesign, Hampshire, UK). A final reaction mixture (20 μL) included 10 μL PrecisionPLUS mastermix (Primerdesign, Hampshire, UK), 5 μL of DNA, 0.4 μM of COG II primers and 0.1 μM of Ring 2 P probe or 1 μL of the IC primer/probe mix (Primerdesign, Hampshire, UK). The thermal profile for the two-step assay consisted of 95°C for 2 minutes followed by 40 cycles of 95°C for 10 seconds and 60°C for 60 seconds**.** Probe fluorescence, for the one and two-step assays, was acquired on the green channel during annealing and extension.

For the plasmid template, a single standard curve was run with the one-step qPCR assay, 6 replicates, and three independent standard curves with the two-step qPCR assay, 20 replicates overall.

For the ssRNA template, three independent standard curves were run with the one-step qPCR assay, 14 replicates overall. A one-step (no RT control) and two-step qPCR assay were performed on each dilution to detect the presence of residual DNA.

For each respective template, the mean Ct value and calculated copy number were averaged for each 10-fold dilution, the slope and y-intercept could then be calculated by linear regression analysis.

Table 2.3 Primer and probe oligonucleotides.

Primer/ Probe	Sequence (5'-3')	GC[a] (%)	Tm[b] (°C)	Amplicon size (bp)
RNA SC F	**TAATACGACTCACTATA**GAGGGTGAATGGATTTTT	34	59.7	181
RNA SC R	AGGCCGGCGGCACCATCATTAGATGG	62	64.3	
GADPH F	GAACGGGAAGCTTGTCATCA	50	51.8	209
GADPH R	ATCGCCCCACTTGATTTTG	47	48.9	
GADPH P	CCCATCACCATCTTCCAGGAGCGAGA[c]	58	62.7	
COG 2 F	CARGARBCNATGTTYAGRTGGATGAG	35-58	53.2-62.7	97
COG 2 R	TCGACGCCATCTTCATTCACA	48	52.4	
Ring 2 P	TGGGAGGGCGATCGCAATCT[c]	60	55.9	

F:Forward, R:Reverse, P:Probe, References – GADPH set: (Miyamoto et al. 2011), COG set: (Kageyama et al. 2003).
[a]Guanidine / Cytosine content.
[b] Melting temperature (20 μM - Basic).
[c] FAM reporter dye coupled to the 5'/TAMRA quencher dye coupled to the 3' end of the oligonucleotide.

2.3 Results

2.3.1 The commercial EIC is suitable to assess the efficiency of clinical stool extraction

To use the EIC in experiments, it was necessary to measure the consistency of template detection during clinical sample extraction, with and without the presence of GII Nov. The one-step qPCR assay measured the IC to be suitable for monitoring the recovery of RNA during extraction in the presence or absence of virus (Table 2.4).

Table 2.4 EIC validation between a HuNoV positive and negative sample.

	Ct value	
Sample	HuNoV	IC
A	20.03	30.27
N	Negative	30.8

2.3.2 Duplex quantification of the commercial IC and GII HuNoV influences virus quantification efficiency

The effect of an IC and GII HuNoV duplex was evaluated by one-step qPCR of the virus in the presence and absence of IC amplification. Relative quantification of the virus in the presence of the spiked RNA template was inhibitory toward GII HuNoV quantification in most reactions (7/8) (Table 2.5). The mean increase in the HuNoV Ct value was 1.07 and ranged from 0.1 to 3.0; therefore, assays hereafter would amplify both in separate reactions (Table 2.5).

Table 2.5 GII HuNoV quantification in single reactions and an EIC multiplex.

Sample	Experiment	HuNoV Ct value without IC	HuNoV Ct value with IC
A	1	18.76	18.93
	2	20.19	21.65
	3	19.44	21.17
	4	19.36	19.98
B	1	19.64	18.75
	2	19.36	19.46
	3	18.23	18.66
	4	18.62	21.62

2.3.3 Weight normalization of faecal samples provides greater consistency than an endogenous housekeeping gene

Two methods of stool sample normalisation were compared; these were weighing the amount of stool used for extraction or quantifying the endogenous reference gene, glyceraldehyde-3-phosphate dehydrogenase (GADPH), in parallel one-step qPCR reactions.

The average $\log_{10}$ of viral load for two different sets of repeated sample preparations, deduced from the validated RNA standard (Figure 2.2), were 9.7±0.3 for A and 9.9±0.25 for B (Table 2.6). These results demonstrate the reproducibility of HuNoV sample preparation, nucleic acid extraction and one-step qPCR. Moreover, the results from one-step quantification in the same sample were indicative of a homogenous distribution of virus in the HuNoV positive stool samples tested.

Table 2.6 Weight normalisation of faecal suspensions from two samples.

Sample	Experiment	Weight (g)	HuNoV Ct value	Viral load (genome copies/g)	Log_{10} viral load
A	1	0.05	18.76	1.1×10^{10}	10.1
	2	0.09	20.19	0.26×10^{10}	9.4
	3	0.10	19.44	0.37×10^{10}	9.6
	4	0.11	19.36	0.36×10^{10}	9.6
B	1	0.02	19.64	1.6×10^{10}	10.2
	2	0.10	19.36	0.36×10^{10}	9.6
	3	0.09	18.23	0.89×10^{10}	9.9
	4	0.11	18.62	0.57×10^{10}	9.8

Detection of the reference gene, GADPH, in stool samples, with or without GII HuNoV, was consistently low between experiments; ranging from 35.99-38.35 Ct values (Table 2.7). Quantification of the reference gene was more variable with standard deviations ranging from 0.05-1.67, in comparison to GII HuNoV in weight normalised samples, 0.59-0.65.

Table 2.7 Reference gene normalisation.

Sample	Weight (g)	GADPH Ct value	HuNoV Ct value
A	0.05	Negative	18.78
	0.04	36.29	19.12
B	0.13	38.35	25.94
	0.11	35.99	26.1
C	0.07	36.36	Negative
	0.05	36.29	Negative

2.3.4 A HuNoV primer specific two-step qPCR approach showed the greatest sensitivity for GII HuNoV detection

To further develop an assay for GII HuNoV quantification from clinical samples, experiments were undertaken to amplify a dilution series of the virus in sample A, using one and two-step qPCR with random or HuNoV-specific RT reactions. A two-step qPCR with a virus targeting RT step was the most sensitive of all the PCR assays tested (Table 2.8). In contrast a randomly primed RT, that shouldn't discriminate between sequences prior to RNA conversion, was the least sensitive for detection of GII HuNoV of all three assays performed. The one-step qPCR assay (RT and PCR combined) showed an intermediate level of sensitivity, but at the highest dilution factor the reaction was in the limits of detection for the virus (Table 2.8).

Table 2.8 GII HuNoV positive stool sample RT-PCR results using different PCR/priming strategies.

Sample	PCR reaction	Dilution factor	HuNoV Ct value	Random hexamer Ct value
A	One-step	10^{0}	19.25	
		10^{-1}	21.33	
		10^{-2}	24.24	
		10^{-3}	28.39	
		10^{-4}	37.90	
	Two-step	10^{0}	17.20	21.57
		10^{-1}	20.52	24.58
		10^{-2}	22.75	28.16
		10^{-3}	27.58	31.72
		10^{-4}	29.82	34.56
		10^{-5}	33.28	37.37

To examine whether differences in sensitivity could be observed in a different type of template these experiments were repeated to quantify the *in vitro* transcribed RNA standard curve. The *in vitro* transcribed curve was assayed in triplicate from 3×10^6 to 3×10^3 copies, and the sensitivity increased consecutively from the two-step assay with a randomly primed RT, to the one-step assay and further to the targeted two-step qPCR, with the exceptions of 3×10^4 and 3×10^3 copies for this assay (Figure 2.2). From 3×10^4 to 3×10^3 copies the targeted two step lost the ability to detect GII HuNoV, whilst the two-step with a randomly primed RT became increasingly variable (Figure 2.2). In contrast, the one step assay did not become undetectable from 3×10^4 to 3×10^3 copies and was much less variable within this range than a two-step with a randomly primed RT (Figure 2.2).

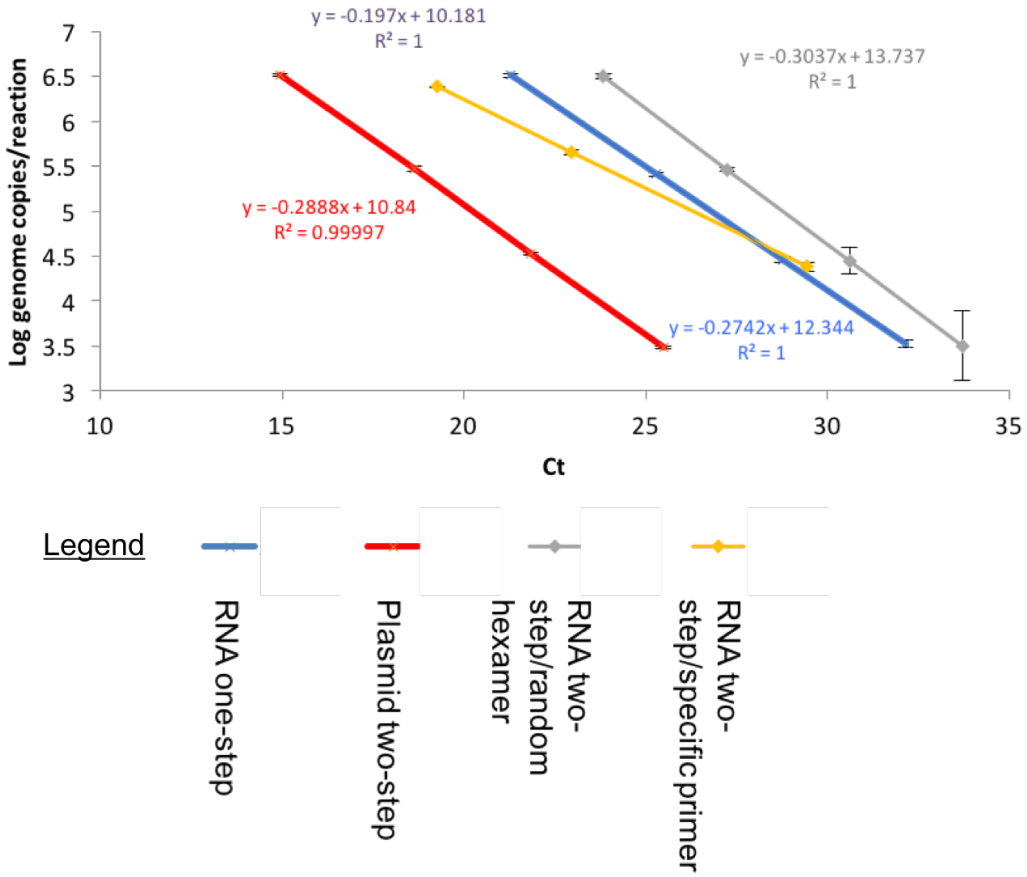


Figure 2.2 Standard curves generated in one and two-step qPCR using two template preparation.

2.3.5 Absolute quantification with a plasmid standard can underestimate viral load

Comparison of the two different standards, *in vitro* transcribed RNA and plasmid, after validation highlighted that the RNA standard generated a higher Ct value at an identical predicted copy number to the DNA (Figure 2.2).

To investigate this, standardized and published reference materials, lenticules, containing a geometric mean of 4.04 $\log_{10}$ GII HuNoV genome copies/disc were diluted and quantified with the RNA and DNA standards. Lenticule dilutions 10^{1},10^{-1} and 10^{-2} were estimated to be at $\log_{10}$ copy numbers 3.39±0.17, 2.78±0.14 and

1.57±0.089 against the DNA standard, whereas the RNA template estimated greater viral titres of 5.31±0.17, 4.72±0.13 and 3.57± 0.082 (Figure 2.3). The batch of lenticules measured had an expected range of 1.65 to 4.46, which was in close agreement with the plasmid standard, and hence the HuNoVs in the disc were quantified with a DNA standard (Figure 2.3 - A). If quantification of the reference material was performed with the DNA standard, the log genome copy number appeared to be underestimated when compared to the RNA standard (Figure 2.3 - B).

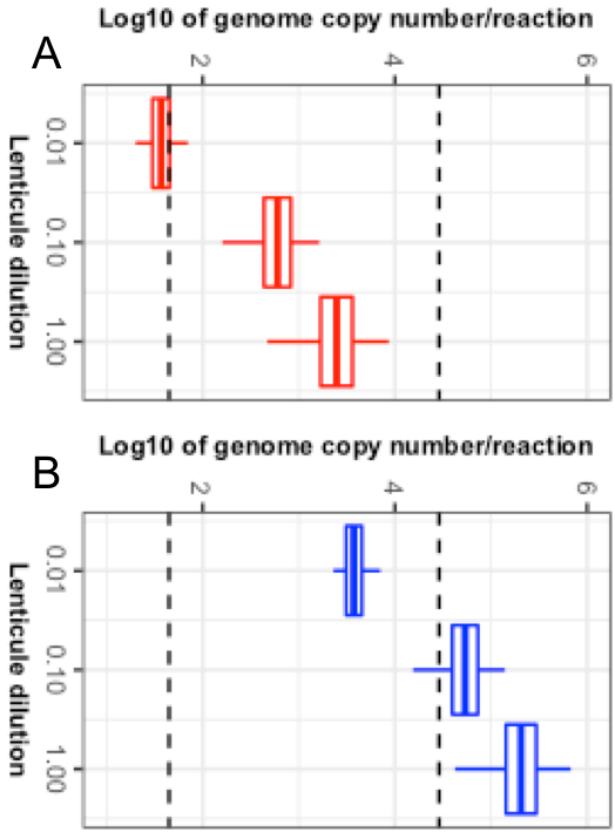


Figure 2.3 Quantification of lenticule reference material with different standards A) Plasmid B) *in vitro* transcribed RNA (Dashed lines = Upper and lower expected range of the lenticule batch).

2.3.6 Application of the absolute quantification of GII HuNoV method to quantify HuNoV present in samples collected longitudinally from the acute phase of infection to convalescence

Longitudinal samples from three different patients (D, E and F) were weight normalised and quantified in triplicate by one-step qPCR, in parallel with the two standards (Figure 2.4). Where samples were available from the day of symptom onset, peak viral load was observed approximately 24 hours from onset, with decreasing titres subsequently (Figure 2.4). The data shows that this method is highly reproducible, and the use of a DNA standard for HuNoV load estimation consistently underestimates the amount of virus present in the clinical samples by nearly 3 logs of magnitude.

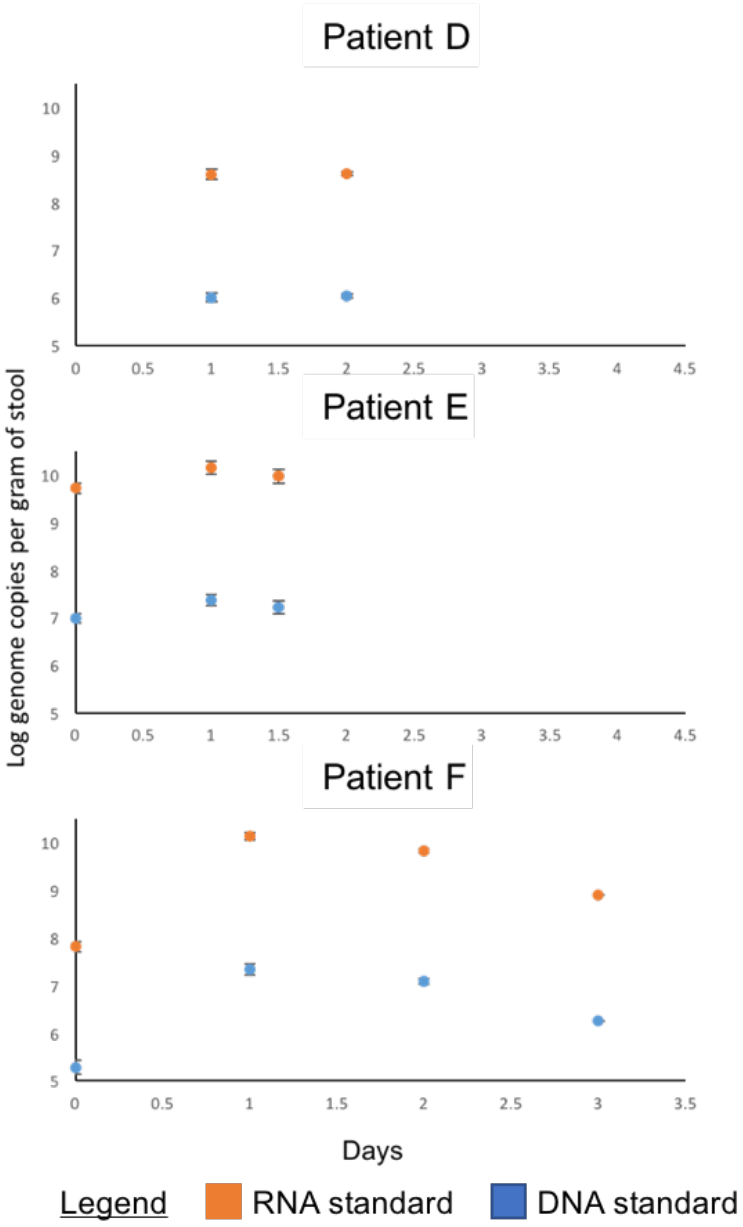


Figure 2.4 Quantification of longitudinal GII HuNoV samples with the optimised method.

2.4 Discussion

In this series of investigations, methods for consistent HuNoV quantification in clinical samples have been optimized.

Firstly, a commercially available EIC to monitor the extraction efficiency of viral RNA from stool was validated. Multiplexing the target virus-specific and EIC PCRs

together resulted in increasing Ct values for the target PCR in most reactions. This inhibitory effect could be due to competition between the two templates; directly for primer binding or indirectly as reagents become limiting in the PCR reaction. The concept of a bias toward the amplification of one template over another is documented in the literature, and can be split into PCR selection and drift (Polz and Cavanaugh 1998; Suzuki and Giovannoni 1996; Wagner et al. 1994). Overall, the experimental results would suggest the presence of PCR selection rather than drift, as the inhibitory effect was reproducible. However, PCR selection is multifactorial; and can be influenced by primer GC content, template length, template accessibility and the number of primer binding sites in a template (Polz and Cavanaugh 1998).

Secondly, normalization and quantification by stool weight was established as the most reliable method when compared to the use of a house keeping gene, GADPH. The use of housekeeping gene amplification for relative quantification is common practice in clinical virology, where the method has been validated for tissue samples and blood primarily (Dong and Wang 2012; Gra et al. 2003; Jabs et al. 2001; Kuchipudi et al. 2012). However, the validity of this approach for use with stool samples is far less clear. Stool samples are heterogeneous, and will contain varying proportions of host cell material, microorganisms (bacteria, virus and potentially protozoa), partially digested foods (of animal or plant origin, and hence may also contain pathogens or symbionts of those organisms). Therefore, accurate normalization of quantification is severely challenged by the enormous diversity from sample to sample not only depending on the individual but also on the severity of the symptoms when samples are being investigated for GE, the duration

of symptoms and the degree of injury to the gut. Although some reports have suggested that GADPH or other human housekeeping genes can be applied for relative quantification purposes (Miyamoto et al. 2011); the data presented here suggest that this is unreliable due to the levels detected, when present, being low and close to the assay's limit of detection. Another challenge is that viruses form clumps either because they are attached to cell debris or linked by antibodies, particularly as infection and recovery progress. The experiments presented here clearly indicate that normalization by stool weight is a reproducible approach in acute HuNoV infection, and furthermore, that in this environment the virus is represented homogenously, hence making norovirus quantification using this approach robust and reproducible.

Thirdly, for HuNoV RNA amplification, a two-step RT-PCR approach using HuNoV specific primers in the RT step was established as the most sensitive approach for detection and quantification. By separating the RT conversion step and DNA amplification, both reactions can be optimized without limiting substrate or reaction conditions. In contrast, a two-step with random hexamer priming in the RT reaction reduced sensitivity although this was within the instrumental error. Therefore, a random RT could be viable in obtaining cDNA across the breadth of a sample without having a significant influence on detection sensitivity. The advantage of the latter approach is that the cDNA produced could be used for the detection of multiple targets that may be other RNA viruses, transcripts of interest, and other regions of the genome of interest. A random RT of a HuNoV positive clinical sample could be used to amplify the major capsid gene that may be

necessary for virus characterization and for outbreak investigation and tracking (Iturriza-Gomara et al. 1999). Furthermore, in sequencing projects it is important to analyse an even distribution of a virus population, without bias toward a single genotype. HuNoV, an RNA virus, can mutate readily in a cloud of quasispecies leading to variants and eventually generation of novel genotypes (Domingo et al. 1985); the use of random priming can maintain the detection of diversity with a reduction in bias in diagnostics and virus characterization assays used in reference laboratories and research.

Finally, development and the use of an RNA standard curve demonstrated quantitative methods using DNA standard curves underestimate viral load, which are widely used in HuNoV detection (https://www.iso.org/standard/65681.html). This is an important observation, and is particularly pertinent and relevant to the quantification of virus in foods and the environment, where assumptions may be made on the risk to users and consumers from contaminated food or water based on viral load measurements.

We have optimised methods for GII HuNoV quantification by qRT-PCR, and applied this methodology to clinical HuNoV samples. First, we have validated a commercially available EIC to monitor the extraction efficiency of viral RNA from stool. Second, stool samples can be normalised by weight more efficiently and reproducibly than with a housekeeping gene (GADPH), for virus quantification. Third, a GII HuNoV specific RT followed by PCR was the most sensitive for viral load quantification in clinical samples.

3 The development of an unbiased capture method for HuNoV

3.1 Introduction

Several nucleic acid sequencing technologies have been invented in the past half-century, which have increased the amount of molecular data available, and further, the applications and different types of analyses that can be performed. The first assays, to determine the order of a nucleic acid sequence, consisted of specific enzymatic or chemical degradation of very short oligonucleotides and product separation by chromatography, electrophoresis or both (2D fractionation), to resolve each fragment (Holley et al. 1964; Sanger et al. 1965). Over a decade later, the concept of sequencing by DNA polymerase extension was introduced, which required individual radiolabeled nucleotides to be added sequentially for step wise extension, and chromatographic separation of each product (Padmanabhan et al. 1972). Thereafter, Sanger published a methodology to determine 50 nucleotides, the 'Plus-Minus system', by combining 5' to 3' nucleotide addition of DNA polymerase and 3' to 5' cleavage of T4 polymerase, which immediately predates widely used 1st generation sequencing technologies (Sanger and Coulson 1975).

In 1977, two 1st generation sequencing assays were published, which detected deoxyribonucleic acid (DNA) strand fragments radiolabeled on the phosphate group at the 5' end by electrophoresis. Firstly, the 'Chemical sequencing' method (50-100 nucleotides) which chemically degraded a specific nucleotide base, which was followed by strand scission, to produce fragments at each position of the DNA strand (Figure 3.1 - A) (Maxam and Gilbert 1977). Secondly, 'Sanger method' (~1000

nucleotides) which combined primer directed DNA replication with radiolabelled dideoxynucleotides (ddNTPs) to limit strand extension at each position (Figure 3.1 - B) (Sanger et al. 1977).

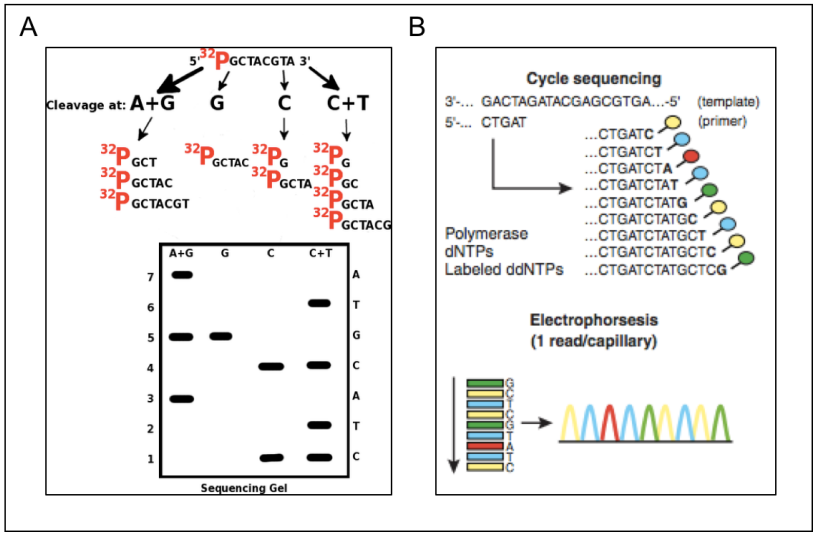


Figure 3.1 The process of (A) Maxam-Gilbert sequencing (https://commons.wikimedia.org) and (B) Sanger sequencing to discern nucleotide sequences (https://www.slideshare.net/ueb52/introduction-to-next-generation-sequencing-v2).

The 'Sanger method' became the first widely used 1st generation sequencing technology, and further improvements to this methodology such as the use of fluorophores and the detection of fluorescence by a laser allowed the sequencing to take place in a single reaction (Figure 3.1 - B). Sequencing then became accessible to many laboratories through reduced economic costs, due to vastly reduced running times, and simplified protocols after the introduction of capillary sequencing machines coupled with the ability to obtain sequence for increasingly longer fragments. This marked the beginning of the "genomic revolution", the exponential increase in DNA sequence data availability, and the routine application of sequencing to a wide range of research and medicine areas. These improvements

have been carried forward to subsequent sequencing technologies (Hunkapiller et al. 1991). The sequencing technologies currently available can be separated into 1st (e.g Sanger sequencing), 2nd (Massively parallel sequencing platforms) and 3rd (Single molecule sequencing platforms) generation technologies (Heather and Chain 2016).

3.1.1 Massively parallel sequencing

Massively parallel sequencing (MPS) is a process, provided currently by multiple platforms, that allows sequencing of all DNA in a sample with high redundancy. The process can be split into three stages: library preparation, sequencing and data analysis.

At the library preparation stage, nucleic acids are fragmented and indexed. Fragmentation can be achieved mechanically (nebulization or hydrodynamic shearing), enzymatically (transposase or nuclease) or chemically (heating with a divalent cation). Individual fragments are then labelled with a unique index sequence to allow for sample pooling in the latter stages. At this point, if necessary, the library can be amplified by polymerase chain reaction (PCR), depending upon the amount of template available.

At the sequencing stage, if the library is not already immobilized (such as via emulsion PCR), the library is bound to a solid surface by oligonucleotides complementary to the index or by the presence of an immobilized polymerase. The different platforms available can then be separated into the following categories based on sequencing chemistry: cyclic reversible termination (CRT), single-

nucleotide addition (SNA), sequencing by ligation (SBL) and real-time sequencing (RTS) (Table 3.1).

3.1.2 The Massively Parallel Sequencing platforms

Different types of MPS platform exist; each varies in sample preparation, throughput, read length, read accuracy and the amount of data produced (Table 3.1). The aims of a project should dictate the choice of platform; for example, a project attempting to identify quasispecies in a population of small genomes would not require long reads but would need a high base call accuracy.

Table 3.1 Types of MPS platforms and their specifications.

Manufacturer	Sequencing chemistry	Approximate read length (bases)	Approximate maximum amount of data per run	Strength	Weakness	Reference Study or Review
Illumina (HiSeq/MiSeq)	CRT	36-600	540-15,000 Mb	High throughput with a low error rate	Sequencing AT-GC rich regions is a challenge	(Radford et al. 2012)
Thermo Fisher Scientific (SOLiD)	SNA	200-400	30-2,000 Mb	Fastest throughput and shortest run time	Poor performance with homopolymers	
Roche (454)	SNA	400-700	450-700 Mb	Long reads	Difficult to distinguish signal for homopolymers of 6 nucleotides and above	
Pacific Biosciences (PacBio)	RTS	0-60,000	500-1,000 Mb	Long reads	Accuracy 83% on one run (can increase with repeated sequencing runs)	(Metzker 2010)
Thermo Fisher Scientific (Ion Torrent)	SBL	50-75	80,000-320,000 Mb	Positions sequenced twice – high accuracy	Instrument has difficulty sequencing AT-GC rich regions	
Oxford Nanopore Technologies (MinION)	RTS	10,000-100,000	40,000 Mb	Longest reads	Difficulty sequencing GC rich regions/High error rate	(Laver et al. 2015)

Approximate read length and maximum amount of data per run obtained from manufacturer's website (2016)

3.1.2.1 Bridge amplification (2nd generation)

Bridge amplification is a sequencing process provided by the HiSeq and MiSeq platforms (Illumina). Fragmented nucleic acids are bound to a lawn of oligonucleotides on a flow cell via adapters, and the complementary sequence generated. Each adapter, present on the 5' and 3' can then form a bridge with their corresponding adapter on the flow cell for clonal amplification. Once clusters are generated the 3' ends are blocked and, firstly, forward strands are sequenced with

fluorescently labeled reversible terminators, followed by the reverse strand in the next stage (Figure 3.2 - **F**) (Bentley et al. 2008).

3.1.2.2 Oligonucleotide ligation and detection or SOLiD sequencing (2nd generation)

Oligonucleotide ligation and detection sequencing requires nucleic acid fragments to be bound to beads covered with adapter sequence, which are then clonally amplified and further aggregated to larger polystyrene beads for fixation to a glass slide. Sixteen different fluorescently labelled primers are introduced (octamers), corresponding to four colours, and successively ligated via the first two nucleotides. The latter six nucleotides of the primer are universal, of which the final three are bound to the fluorophore, and chemically cleaved after laser excitation. However, as two nucleotides on the probe dictate binding, and the following three are universal bases which are not cleaved, the sequence must be offset by 1-5 positions to obtain the complete sequence data (Figure 3.2 - **B**) (McKernan et al. 2009).

3.1.2.3 Pyrosequencing (2nd generation)

The concept of measuring the production of inorganic phosphate in biological reactions was published in 1985 (Nyren 1987). Nyren and colleagues demonstrated inorganic phosphate could be sequentially converted to adenine triphosphate (ATP), by ATP sulfurylase, and then used as a substrate for bioluminescence, by luciferase (Nyren 1987). A platform, the 454 instrument, was then manufactured (Roche) that could measure the release of organic phosphate after the correct incorporation of a nucleotide by DNA polymerase, by cycling through the four different options (Figure 3.2 - **E**) (Margulies et al. 2005).

3.1.2.4 Semiconductor sequencing (2nd generation)

In semiconductor sequencing, provided by the Iontorrent platform (Life technology), nucleic acid fragments are bound to beads and clonally amplified. After amplification, each bead occupies a single well in a chip and an individual nucleotide is added for incorporation by DNA polymerase; if the correct nucleotide is added a proton is released which can be detected as a pH change (Figure 3.2 - **D**) (Rothberg et al. 2011).

3.1.2.5 Single molecule real-time sequencing (3rd generation)

Single molecule real-time sequencing, provided by the PacBio platform (Pacific Biosciences), does not require clonal amplification. This platform detects strand extension as the fluorophore present on labeled nucleotides is cleaved by DNA polymerase in real-time (Figure 3.2 - **C**) (Braslavsky et al. 2003).

3.1.2.6 Direct strand sequencing - Nanopore (3rd generation)

In a similar manner to the PacBio, nucleic acids prepared for direct strand sequencing are not clonally amplified, and can be performed with the MinION platform (Nanopore technology). Direct strand sequencing requires an electrophoretic gradient separated by a bilayer which contains pores, such as α-hemolysin. A strand of ribonucleic acid (RNA) or DNA diffuses through the pore, down the gradient, and each base blocks an electric current running across the bilayer to produce a characteristic signal disruption in the electric current (Figure 3.2 - **A**) (Kasianowicz et al. 1996).

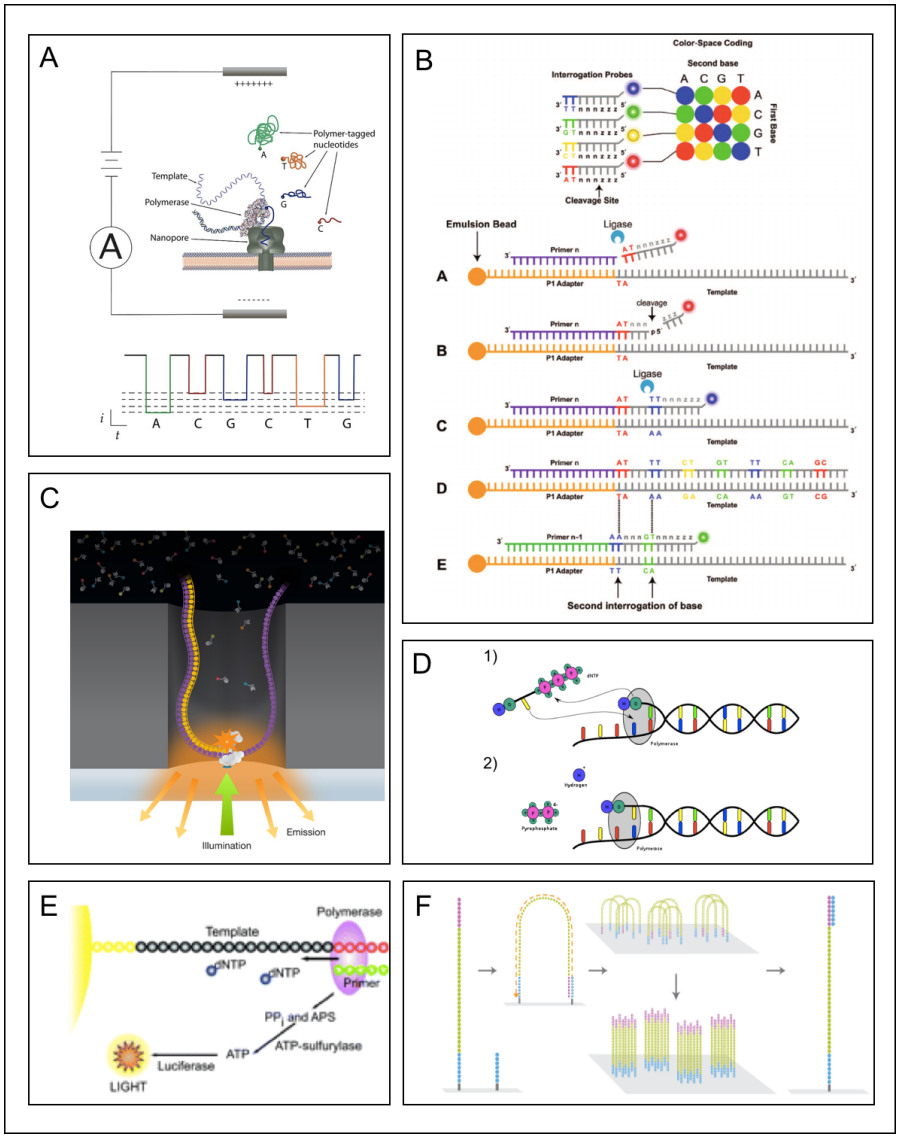


Figure 3.2 The types of sequencing chemistries currently available – (A) Direct strand sequencing (https://commons.wikimedia.org), (B) Oligonucleotide ligation and detection sequencing (Voelkerding et al. 2009), (C) Single molecule real time sequencing (http://data-science-sequencing.github.io/lectures/lecture6/), (D) Semi-conductor sequencing (https://commons.wikimedia.org), (E) Pyrosequencing (Voelkerding et al. 2009), (F) Bridge amplification (https://www.slideshare.net/AndorKiss/new-technologies-at-the-center-for-bioinformatics-functional-genomics-at-miami-university).

3.1.3 Massively parallel sequencing vs Sanger sequencing

Traditional Sanger sequencing is limited by the requirement of targeted amplification to initiate sequencing and the small volume of sequence data produced in comparison to MPS methods. A Sanger sequencer can detect a second

allele in a mixed population at a frequency of >20% at best (Bar-Eli et al. 1989). In contrast, an Illumina platform can generate sufficient depth of sequence data to reliably call minority variants present with a frequency of > 1% and above (Thys et al. 2015). Furthermore, prior to MPS, amplification of the target by PCR is not an absolute requirement, and hence selection and bias during amplification can be reduced (Wagner et al. 1994).

3.1.4 Massively parallel sequencing in virology

The introduction of MPS platforms has enabled *de novo* unbiased sequencing, and is considered an ideal tool for investigations of metagenomes, viromes and pathogen discovery. Furthermore, the ability of MPS to detect minority species has provided great potential to survey virus diversity. An increase in the resolution of minority species has meant virus evolution can be measured within or between hosts in transmission studies, and has allowed for the early detection of emergent antiviral resistance.

However, apart from direct strand sequencing (Kasianowicz et al. 1996), MPS platforms are limited by the requirement of a reverse transcription (RT) step for library preparation and further downstream processing of RNA molecules. Therefore, investigations into viruses with RNA genomes require an additional RT conversion step, which can vary in efficiency. The efficiency of an RT step is influenced by the total amount, or fragment size, of nucleic acids in a sample, and is a significant contributor to sequence error if coupled with a high-fidelity polymerase in PCR (Arezi and Hogrefe 2007; Nolan et al. 2006; Ståhlberg et al. 2004).

3.1.4.1 Viral metagenomics

Viral metagenomics is the indiscriminate examination of all the nucleic acids present in a sample, often following a universal virus particle enrichment protocol, to identify undiscovered viruses and novel roles for those viruses already identified. Notable viral metagenomic discoveries, by MPS, include the discovery of Schmallenberg virus in cattle (Hoffmann et al. 2012a) or the strong association of IAPV (Israeli Acute Paralysis Virus) as a marker for Colony Collapse disorder in Honey Bees (Cox-Foster et al. 2007).

3.1.4.2 Virus surveillance

Virus surveillance is commonly applied in epidemiological studies, to understand transmission patterns on a local and global scale, or in a clinical setting for the detection of antiviral resistance in hosts undergoing therapy.

MPS can be applied to environmental or clinical samples to track the transmission of a mutant virus that exists in a virus population at low frequency, but becomes the dominant haplotype after a bottleneck transmission event (Bull and White 2011). Once the source of an outbreak or zoonotic transmission event is mapped the information can be used to implement preventative measures. The ability to identify the source of an emergent variant has been important when tracking cross species barrier jumps of Influenza A, which could lead to the introduction of a novel, more pathogenic, virus (Croville et al. 2012).

In a clinical setting, MPS has been used to identify the presence of underlying antiviral resistant variants in a virus population that could be selected for during therapy. In the literature, the 454 (Roche) MPS platform has been used to detect

the appearance of antiviral resistant variants early in chronic Hepatits C virus (HCV) infection, or to measure the population diversity (and receptor usage) in HIV positive hosts to better inform on the most effective therapeutic option (Abbate et al. 2011; Newman et al. 2013).

3.1.5 HuNoV: Partial and whole genome sequencing

The field of HuNoV molecular epidemiology has benefited from the development of RT-PCR methodologies to detect and sequence short regions of the genome. Diagnostic RT-PCR assays were meant to be highly sensitive for detecting diverse HuNoV for national surveillance, and hence targeted the most conserved region of the genome (Ando et al. 1995; Green et al. 1995; Vinje and Koopmans 1996; Wang et al. 1994).

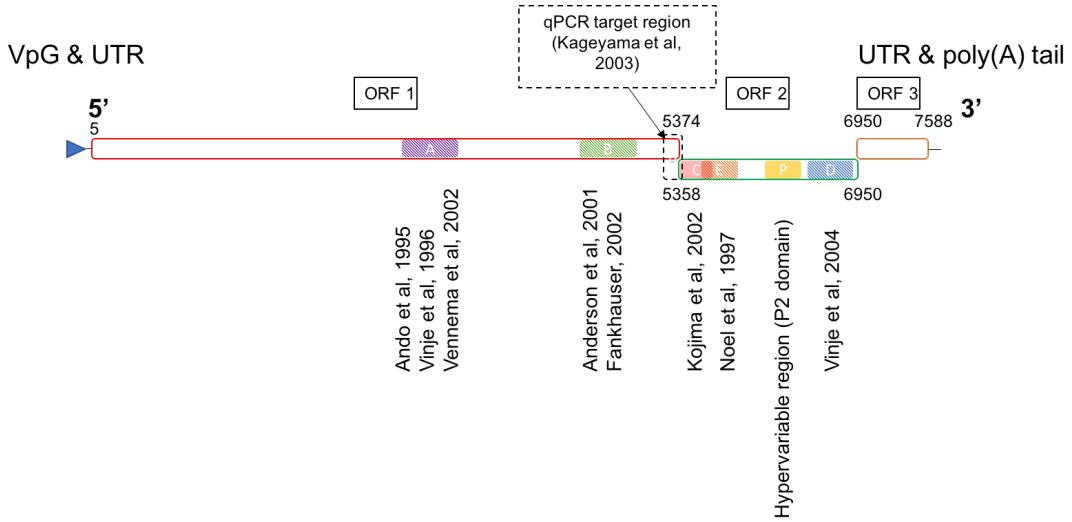


Figure 3.3 HuNoV primer regions used for diagnostic and surveillance RT-PCRs mapped to the Norwalk virus genome (M87661).

Table 3.2 HuNoV diagnostic primer positions mapped to the Norwalk virus genome (M87661).

Primer sets	Region	Positions	Technique	Classification level	Reference
SR33, SR46, SR48 SR50, SR52	A	5362-5683	Conventional PCR	Genogroup	Ando et al, 1995 Vinje et al,1996 Vennema et al, 2002
Norwalk-like virus 431 to 434	B	5117-5285	Conventional PCR & Sanger sequencing	Genotype	Anderson et al, 2001 Fankhauser, 2002
G1F1, G1R1, G1SKF, G1SKR	C	5342-5671	Conventional PCR	Genogroup	Kojima et al, 2002
Cap A, Cap B2, Cap B1	D	6738-6914	Conventional PCR & Sanger sequencing	Genotype	Vinje et al, 2004
Mon381, Mon383	E	5362-5683	Conventional PCR & Sanger sequencing	Genotype	Noel et al, 1997
Several sets exist	P2 domain	6192-6575	Conventional PCR & Sanger sequencing	Genotype/with sequencing genotype variant strains	Allen et al, 2008
COGF1, COGR1 COGF2, COFR2	ORF1/2 junction	5291-5375	qPCR	Genogroup	Kageyama et al, 2003

To characterise HuNoV strains, amplification and sequencing of more diverse regions of the genome has been routinely used in reference laboratories. Partial sequencing of regions of the virus capsid and RNA dependent RNA polymerase (RdRp) have been used for virus characterisation, strain surveillance and

transmission studies. Furthermore, studies in which regions of the ORF1 as well as the ORF2 have often identified potential recombinant strains (Waters et al. 2007). However, such findings need to be confirmed through sequencing of contiguous sequence spanning both ORFs.

3.1.6 Massively Parallel Sequencing of HuNoV and enrichment

MPS provides the opportunity to obtain information from the entire genome, the ability to detect minority variant strains and virus mixtures in dual or multiple infections. HuNoV cannot be readily propagated in cell culture (Duizer 2004), and therefore virus detection and characterisation is typically done directly from clinical samples (stool). Whilst this is advantageous, as it precludes virus selection, several disadvantages exist in the application of MPS directly to nucleic acids extracted from stool samples.

HuNoV RNA will represent a very small proportion of the total nucleic acids present in a stool sample (Nakamura et al. 2009). Therefore, prior to library preparation it is necessary to isolate or at least enrich the HuNoV RNA against host, bacterial and other viral nucleic acids to ensure sequencing of the genome of interest. A handful of HuNoV MPS studies exist in the literature; in these studies, the availability of the virus genome for sequencing has been maximized by PCR, specific or non-specific virus RNA enrichment or a combination.

3.1.6.1 Ribosomal RNA depletion

A complex microbiota is a significant component of the gut, and it is shed as a component of stool alongside host cells by movement of the digestive tract (Eckburg et al. 2005). Bacterial and host cell ribosomes are comprised of protein

and non-coding RNA. Therefore, after total nucleic acid extraction of a stool sample, a large proportion of the RNA isolated will inevitably be ribosomal. To increase the representation of HuNoV RNA, oligonucleotide probes have been adapted to bind the non-coding RNA, and then double stranded nucleic acids can then be removed with magnetic beads (Bavelaar et al. 2015; Hasing et al. 2016). The removal of excess non-target RNA allows for enrichment of the viral RNA of interest, and enhances virus sequencing.

3.1.6.2 Polyadenylated tail selection

Polyadenylated tail selection has commonly been applied to separate non-coding RNA from mRNA in transcriptomics, prior to RNA-seq (Wilhelm and Landry 2009). The presence of a polyadenylated tail at the 3' end is a structural feature of the HuNoV genome. Therefore, the methodologies used to select for messenger RNA (mRNA) from other nucleic acid extracts can be applied to the HuNoV genome. Polyadenylated RNAs can be selected for, from the total extract, by using d(T) oligonucleotides to physically separate out molecules with a poly-A tail or to initiate RT (Batty et al. 2013; Fonager et al. 2017; Wong et al. 2013).

3.1.6.3 RNA baits

A library of RNA baits is a collection of 120-mer oligonucleotide fragments used to capture complimentary target sequences. A HuNoV-specific library is sufficiently diverse to capture multiple genotypes prior to MPS. This technique has been applied to a panel of 507 stool samples, comprising four and ten different GI and GII HuNoV genotypes, respectively (Brown et al. 2016). The RNA bait library could generate complete genomes in 89% of the stool samples, with > 80% of genome

coverage from clinical samples containing high and low viral loads (Brown et al. 2016). The intermediate level of specificity of this technique is more advantageous than targeted PCR. However, RNA baits could still generate a bias or would be unable to detect novel HuNoVs or recombinants in contrast to total RNA-sequencing.

3.1.6.4 Physical methods for virus concentration: Ultracentrifugation

Virions in a clinical sample can be separated from other components by size and physico-chemical properties. HuNoV size selection, prior to MPS, has been achieved by sequential filtration, centrifugation and ultracentrifugation through a density barrier (Vega et al. 2014b). Isolation of HuNoV by physical separation can provide a sample fraction of the highest purity, however the process is highly time-consuming and not suitable for high-throughput (Vega et al. 2014b). Moreover, sucrose gradient ultracentrifugation can cause damage to Respiratory syncytial virus (RSV) particles, although HuNoV VLPs have been shown to maintain capsid integrity and HBGA binding after undergoing a similar purification method (Huhti et al. 2010; Trépanier et al. 1981).

3.2 Materials and methods

3.2.1 Optimisation of HuNoV capture from clinical samples

3.2.1.1 Clinical sample preparation

A 10-20% (v/v) stool stock suspension was made up to 40 mL in sterile PBS (Sigma, Dorset, UK) from pooling two high titre clinical samples, due to an insufficient amount of patient stool sample. The stock suspension was diluted (1 in 100) to

correspond to a lower titre clinical sample (Table 3.3). The neat stock and diluted suspension were used to optimise PGM capture and poly-(A) tail enrichment.

A high titre and a low titre stool stock suspension were used to examine the effects of PGM-MB capture and non-specific amplification on sample metagenomics and the frequency of minority variants detected (Table 3.3).

Table 3.3 The stock stool suspensions used to optimise capture and enrichment methods or examine the effects of capture and non-specific PCR on variant calling frequencies.

Sample ID	Sample name	Ct value	Viral copy number/g
15-13	13-IP-01	15.85	2.7×10^{9}
	15-IP-01		
17	17-IP-01	35.31	1.2×10^{4}

A collection of archived stool samples, positive for GII HuNoV, were gifted to this project from a previous study (Table 3.3). Stool samples were obtained from symptomatic individuals at the Royal Liverpool University Hospital (RLUH) in Liverpool. National Health Service (NHS) ethical approval for sample collection was provided by the Research, Development and Innovation department (RD&I) of RLUH, which required patient consent to collection and use of stool for research purposes.

3.2.1.2 Preparation of PGM-MBs

MagnaBind™ carboxyl derivatized beads (Fisher Scientific, Leicestershire, UK) were prepared as described in the manufacturer's instructions, pooled and stored at 4°C. PGM III (Sigma, Dorset, UK) or BSA as control (Sigma, Dorset, UK) were dissolved in conjugation buffer to a concentration of 7.5 mg/mL for the coupling reaction.

3.2.1.3 Optimization of GII HuNoV capture from clinical stool samples

Prior to enrichment, the stool suspension was centrifuged at 5000 rpm for 15 minutes and the PGM-MB stock was reconstituted by pipetting. For each optimization experiment 100 µL of homogenous PGM-MBs was added to 1 mL of capture mixture containing stock stool suspension and sterile PBS (Sigma, Dorset, UK) in a ratio 1:3. The mixture was kept under constant mixing on an SB2 rotator shaker (Fisher Scientific, Leicestershire, UK) for 15 minutes. After mixing the PGM-MBs were isolated by the DynaL magnetic separation (MS) rack (Fisher Scientific, Leicestershire, UK) and washed to homogeneity with 1 mL sterile PBS (Sigma, Dorset, UK) or 0.1 M citric acid-sodium citrate buffer; this was repeated a further two times. The PGM-MBs were then eluted into 60 µL sterile PBS (Sigma, Dorset, UK) and transferred to a sterile microcentrifuge tube.

Various conditions were altered, in triplicate, to optimize PGM-MB virus capture and recovery as follows:

Binding dynamics

- **pH** - Sterile PBS (Sigma, Dorset, UK) in the capture mixture was replaced by 0.1 M citric acid-trisodium citrate buffer at pH 3.6.
- **Incubation time** – The duration of capture incubation was extended to 30 and 60 minutes.
- **Temperature** – The capture mixture was incubated at 4 or 37°C.
- **Receptor saturation** – The ratio of stool suspension was altered (3:1, 1:1, 1:3) or the volume of PGM-MBs added was altered.

Nucleic acid recovery

- **Virus extraction –** Nucleic acids were extracted from the PGM-MB eluate by either heat release, as described previously (Tian et al. 2010), or the guanidium isothiocyanate (Gn) silica method (Boom et al. 1990). Lysis buffer extraction entailed the addition of L6 buffer (Severn Biotech Limited, Worcestershire, UK) or RLT buffer (Qiagen, West Sussex, UK) directly to the eluate, and MS to separate the buffer from the PGM-MBs. The remaining steps of the extraction were performed as described previously (Boom et al. 1990) or according to the manufacturer's instructions. An Internal RNA extraction control (PrimerDesign, Hampshire, UK) was spiked into each sample prior to HuNoV nucleic acid extraction per the manufacturer's instructions.

In PGM-MB virus capture experiments after optimisation, sterile PBS was replaced by 0.1 M citric acid-sodium citrate buffer and reactions were performed at 25 °C for 15 minutes. The captured viral RNA was purified using a Gn silica method (Boom et al. 1990).

3.2.2 Oligo (dT) selection of HuNoV RNA

Nucleic acids were extracted from 0.25 mL of stool stock suspension using the Gn silica method (Boom et al. 1990). Poly-(A) RNA in the extraction eluate was captured using the Poly-(A) Purist Kit (Invitrogen, Renfrewshire, UK). The protocol used for Oligo (dT) selection was performed following the manufacturer's instructions, except for two differences. Firstly, the volume of the extraction eluate was increased to 0.25 mL with DEPC-treated water (Fisher Scientific, Leicestershire,

UK). Secondly, poly-(A) RNA was centrifuged from the column in a volume of 20 µL DEPC-treated water (Fisher Scientific, Leicestershire, UK).

3.2.3 DNase treatment

Prior to library preparation extracted RNA was DNase I (Sigma, Dorset, UK) treated and column purified using the Qiaquick gel extraction kit (Qiagen, West Sussex, UK), following the manufacturer's instructions.

3.2.4 One-step quantitative PCR of GII HuNoV RNA

Quantitative (Taqman) real-time RT-PCR (qPCR) was performed on a Rotor Gene 6000 (Qiagen, West Sussex, UK) using the Rotor-Gene Multiplex RT-PCR Kit (Qiagen, West Sussex, UK). The qPCR targeted the ORF1/2 junction of the GII HuNoV genome to monitor enrichment and library preparation steps (Table 3.2 and Figure 3.3). Each reaction contained 10 µL of 2x Rotor-Gene Multiplex RT-PCR Master Mix (Qiagen, West Sussex, UK), 0.2 µL Rotor-Gene RT Mix (Qiagen, West Sussex, UK), 1 µM COG 2F/COG 2R (Sigma, Dorset, UK) [Sequence: Chapter 2 Table 1], 0.5 µM Ring 2 probe (Sigma, Dorset, UK) [Sequence: Chapter 2 Table 1], 5 µL of RNA sample and DEPC-treated water (Fisher Scientific, Leicestershire, UK) to a total volume of 20 µL. Quantitative real-time RT-PCR cycling conditions were as follows: 50°C for 15 minutes, 95°C for 2 minutes, and then 45 cycles of 95°C for 15 seconds and 56°C for 60 seconds. Fluorescence was detected at the 56°C extension step.

3.2.5 MPS Library preparation

Each Library was prepared using the ScriptSeq v2 RNA-Seq Library Preparation Kit (Illumina, Essex, UK) with one slight modification to the manufacturer's instructions.

HuNoV RNA was denatured at 95°C, and subsequently placed on wet ice, before library preparation. At the amplification stage, Illumina-compatible barcodes replaced the reverse primer to allow for sample multiplexing. Prior to sequencing on the HiSeq2500 Illumina platform at the Centre for Genomic Research, University of Liverpool, library size and concentration was analysed with the 2100 Bioanalyzer (Agilent Technologies) and Qubit dsDNA High Sensitivity assay (Fisher Scientific, Leicestershire, UK), respectively. A sample with an insufficient amount of library present after 12 cycles of non-specific PCR amplification was amplified for a further 10 cycles. If additional amplification was performed, amplicons were purified with AMPure beads (Beckman Coulter, Buckinghamshire, UK) at 0.7x the volume of the total PCR reaction, to minimise the presence of primer dimer.

3.2.6 Data analysis

Perl and R scripts used in the bioinformatics pipeline (Figure 3.4 and Table 3.4) to detect and analyse minority variants are available at:

https://www.dropbox.com/sh/9gw8ub4fyxh2pdf/AAABMJlYyR6oW96jLFeP1uqca?dl=0

3.2.6.1 Consensus mapping and calling minority variants

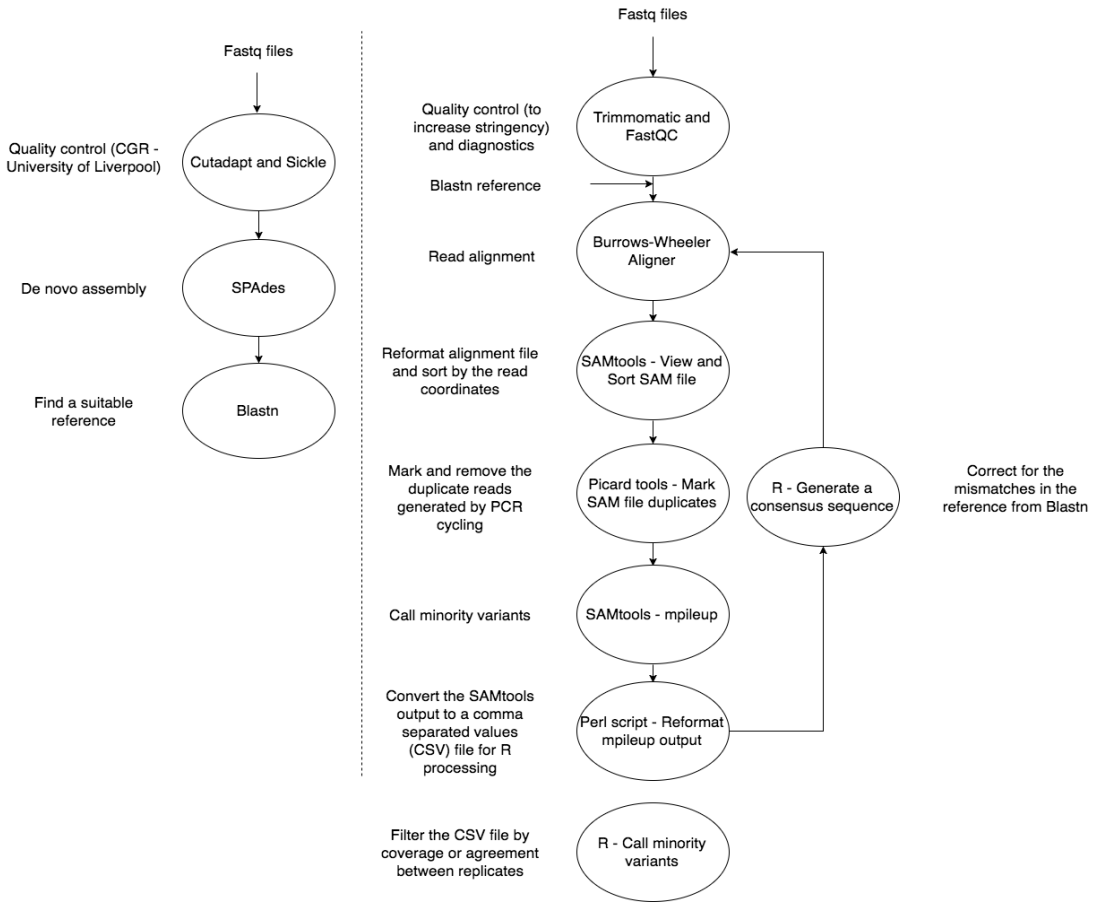


Figure 3.4 Bioinformatics pipeline for generating a consensus sequence and calling minority variants. Perl script to reformat the data is available online at: https://github.com/riverlee/pileup2base.

Table 3.4 References to software used in the consensus generation and minority variant calling pipeline.

Software name	Version	Function	Reference
SPAdes	v.3.5.0	De novo assembly	(Bankevich et al. 2012)
Blastn	2.2.27+	Reference search	(Zhang et al. 2000)
Trimmomatic	0.36	Quality control	(Bolger et al. 2014)
FastQC	V0.10.1	Diagnostics	(Andrews 2010)
Burrows-Wheeler Aligner	0.5.9-r16	Aligner	(Li and Durbin 2009)
SAMtools	0.1.18-r580	Sequence alignment map processing and variant calling	(Li et al. 2009a)
Picard tools	2.1.1	Remove duplicate reads	Available online at: http://broadinstitute.github.io/picard/

3.2.6.2 Metagenomic analysis

To compare the effectiveness of the PGM treatment for HuNoV enrichment metagenomic analytical approaches were used. The Genome Relative Abundance and Average Size (GAAS version 0.17 – available online at: https://sourceforge.net/projects/gaas/files/gaas/GAAS-0.17/) software was used (Angly et al. 2009) for this purpose. Firstly, fastq files were converted to fasta format with Seqtk (https://github.com/lh3/seqtk). Similarities in each fasta file were examined against a RefSeq database (2012) for protozoa, microbes and

viruses. The database was manipulated to add HuNoV consensus sequences from each positive stool sample.

3.2.6.3 Statistical analysis

PGM-MB enrichment optimization data was analysed by a one-way ANOVA.

3.3 Results

3.3.1 PGM-MBs can capture GII HuNoV from clinical samples at high viral loads

The Ct value obtained in NoV specific amplification of the high titre stool suspension before and after PGM-MB capture was not significantly different (< 3 Ct values), despite a slight decrease after capture (Table 3.5).

Table 3.5 Capture of GII HuNoV from clinical samples with PGM-MB.

Sample ID	Ct value	
	No PGM enrichment	PGM enrichment
15-13	25.55	24.2

3.3.2 A low pH environment increases the affinity of PGM-MBs to GII HuNoV

PGM-MB capture of a neat GII HuNoV clinical sample had a lower average Ct when performed in acidic conditions (23.2±1.2) rather than in a neutral pH (24.5±0.27). This effect was significantly more pronounced (32.0 ±1.9 to 27.5 ±0.97) when the sample was diluted 100-fold (Figure 3.5, p = < 0.05).

3.3.2.1 The rate of PGM-MB capture was similar at room or physiological temperature

The PGM-MBs captured GII HuNoV from the neat clinical sample at a similar rate, whether incubated at room temperature (Ct 24.2±0.44) or at physiological temperature (37°C) (Ct 24.7±0.092). However, when the temperature was decreased to 4°C there was a significant loss in virus yield of more than 10-fold (Ct 29.6±0.92), in comparison to the room temperature and 37°C treatments ($p = <0.05$). Temperature of PGM-MB enrichment had a lesser effect on the virus yield obtained with a more dilute sample: Ct 32.7±2.6, Ct 31.4±0.70 and Ct 30.7±0.99, at temperatures 4, 25 or 37°C, respectively (Figure 3.6).

3.3.2.2 Increasing the incubation time during capture did not increase virus capture

If the incubation period of the PGM-MB capture was doubled or quadrupled the GII GII HuNoV yield didn't change (Figure 3.7). Much of GII HuNoV capture took place in the first 15 minutes and did not increase further at 30 or 60-minute incubation periods, this effect could be seen if the clinical sample was neat or dilute (Ct: 26.6±1.1, 26.5±1.2, 26.3±1.1 or 30.8±0.77, 30.9±1.6, 31.1±1.0, respectively).

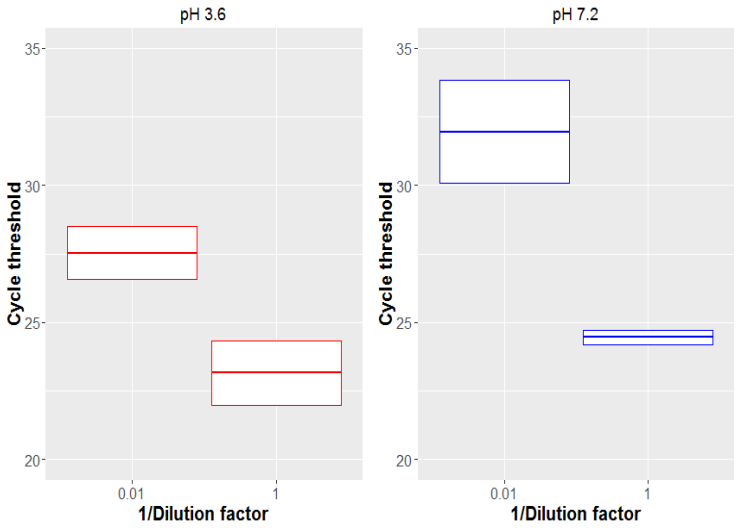


Figure 3.5 The effects of an acidic pH on PGM-MB binding of GII HuNoV.

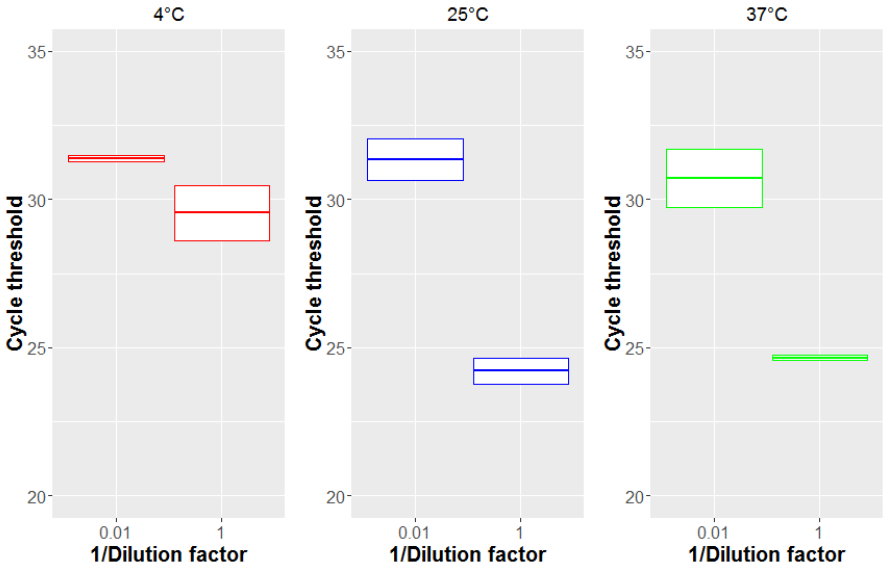


Figure 3.6 The effects of temperature on PGM-MB binding of GII HuNoV.

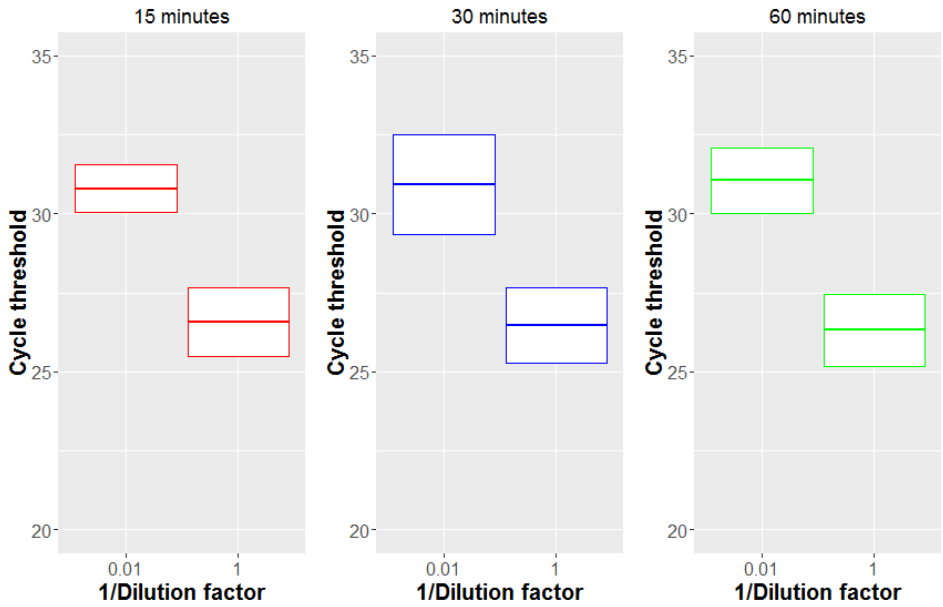


Figure 3.7 The effects of incubation time on PGM-MB binding of GII HuNoV.

3.3.3 Optimization of HuNoV genome extraction after or prior to capture or enrichment

Three different types of nucleic acid extraction methods were assessed to remove the HuNoV genome from the PGM-MB complex (Figure 3.8). There were no major differences in the GII HuNoV Ct value measurements obtained for a column extraction (28.00±0.19), Gn silica extraction (27.6±0.81) or heat release method (27.36±0.59).

A quantitative RT-PCR was performed to examine which extraction method to use prior to oligo (dT) cellulose selection. The Ct value obtained after enrichment was not significantly different when column extraction was applied to the clinical sample (24.69), instead of a Gn silica extraction (25.18).

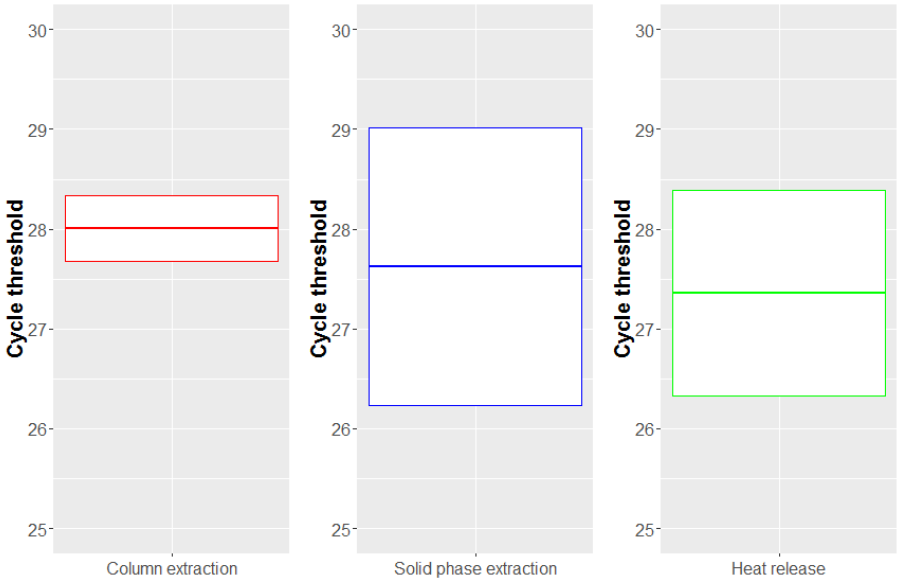


Figure 3.8 HuNoV genome extraction from PGM-MBs.

3.3.4 Enrichment is limited by the buffer used in wash steps, the ratio of PGM-MB to GII HuNoV and the volume of sample available

To examine capture efficiency and whether captured virus is lost during the wash stages, the stool suspension after enrichment, the PGM-MBs and the subsequent washes were extracted and quantified for GII HuNoV.

If a PBS wash buffer was applied, the PGM-MBs captured the GII HuNoV more efficiently, but a loss of viral load was observed at each wash step (Table 3.6). When this experiment was repeated, but the wash was performed with the capture buffer (Citric acid-sodium citrate buffer), less GII HuNoV was lost and in less wash steps (Table 3.6).

Table 3.6 The effects of different wash buffer on PGM-MB capture efficiency measured in triplicate.

Sample A	Enrichment stage	Ct value		
		PBS wash (pH 7.2)	Citric acid-sodium citrate buffer wash (pH 3.6)	
		100 µL PGM-MBs		200 µL PGM-MBs
	Stool suspension after enrichment	29.1±0.39	25.2±0.28	25.6±0.28
	PGM-MBs	26.4	26.4	25.8
	Wash stage I	28.9±0.38	37.4	37.9±2.16
	Wash stage II	27.9±0.27	37.2	Negative
	Wash stage III	29.5±0.78	Negative	34.1

To investigate the binding dynamics of the PGM-MBs further, serial dilutions of a stool suspension were enriched using different sample and PGM-MB volumes. If the volume of the stool suspension was doubled the Ct value tended to decrease proportionally, but there were two exceptions to this trend (Table 3.7). When 100 µL of PGM-MBs was used to enrich 200 µL of diluted suspension the cycle threshold increased in contrast to the corresponding dilutions at 100 µL of stool suspension. These two treatments were the only two examples of an increase in virus yield after halving the volume of PGM-MBs.

If 100 µL of stool suspension was used, doubling the PGM-MB volume caused the enrichment effect to become more pronounced as the sample was diluted (Table 3.7). However, the same effect could not be seen for all dilutions when the volume of stool suspension was doubled.

Table 3.7 Binding dynamics of PGM-MB and GII HuNoV in clinical samples.

Stool suspension volume (µL)	Dilution factor	Ct value		Difference
		50 µL PGM-MBs	100 µL PGM-MBs	
100	1×10^{1}	27.85	26.97	-0.88
	1×10^{-2}	30.9	28.26	-2.64
	1×10^{-4}	34.69	31.39	-3.25
200	1×10^{1}	26.8	25.32	-1.48
	1×10^{-2}	29.03	29.59	0.56
	1×10^{-4}	30.32	32.81	2.49

To identify whether any relationship existed between the amount of virus that could be enriched from a sample and the viral load remaining after enrichment, a panel of 4 different clinical samples were enriched and a comparison was made between the Ct values of the stool suspension, the suspension after enrichment and the PGM-MBs after capture (Table 3.8). For each sample an increase in Ct value was observed from the stool suspension to the amount of virus captured by the PGM-MBs, and a further increase in the Ct value from the captured PGM-MBs to the stool suspension supernatant after capture (Table 3.8). The increase in Ct value from the stool suspension to the PGM-MBs agrees with the binding dynamics investigation (Table 3.7), and likely implies receptor saturation with virus. The Ct value of 12-IP-01 had a much larger Ct value after PGM-MB treatment in comparison to the other samples, which could indicate an inability to access the virus or a loss in titre during the wash steps (Table 3.8).

Table 3.8 The effects of sample variability on PGM-MB virus binding (ND – not detectable).

Sample	Untreated stool suspension Ct value	PGM-MB Capture Ct value	Supernatant (after capture) Ct value
11-IP-01	19.92	22.88	ND
12-IP-01	20.53	28.63	29.15
13-IP-01	24.94	25.42	31.07
13-IP-02	23.3	24.23	30.65

3.3.5 PGM-MB pooling can concentrate HuNoV from clinical stool samples

If the amount of clinical sample enriched with PGM-MBs was increased 3 or 10-fold, and pooled for extraction, the HuNoV yield increased (Table 3.9). An increase of 10-fold corresponded to ~2 decrease in the Ct value (Table 3.9).

Table 3.9 Clinical stool sample pooling strategies.

	No. of sample aliquots pooled by PGM-MBs			
Sample	1	2	3	10
A	22.3	22.43	21.94	20.31

3.3.6 RNA-sequencing of enriched HuNoV clinical samples

The qPCR Ct value decreases in a linear relationship as the virus target sequence increases (Heid Stevens, J., Livak, K.J., and Williams, P.M. 1996), and treatment of the high titre clinical sample with PGM-MB capture and pooling (P10) generated a library with the lowest Ct value, therefore, this library likely contained the most HuNoV sequencing reads (Table 3.10).

Two of the four prepared libraries required further non-specific amplification to have sufficient DNA to be pooled prior to MPS; these were the samples that had no enrichment (NE) and that from a single PGM-MB capture (P1) (Table 3.10). This

meant additional random amplification became an enrichment method to be compared in further experiments. In all the libraries prepared, a lower Ct value correlated with a greater percentage of reads mapped to the GII HuNoV genome and a greater mean coverage (Table 3.10). If less library preparation PCR cycles were performed a lower Ct value meant a higher concentration of DNA (Table 3.10).

Enrichment of the high titre sample generated a complete GII HuNoV genome (PA, P1, P10) after MPS, and for no enrichment (NE) the genome was near complete (99.98%). The mean coverage and percentage of reads mapped was greatest if the sample had undergone poly-A tail selection (PA) or PGM-MB pooling (P10) (Table 3.10). The Ct value obtained after a single PGM-MB capture (P1) was not significantly different from the sample that had not undergone PGM-MB enrichment, with a potential loss of half of the DNA target as indicated by the increase of approximately 1 Ct value (Ct: 17.4 vs 16.57). Also, the PGM-MB (P1) sample had lower mean coverage in comparison to the sample without enrichment (2302.74 vs 3599.58).

Table 3.10 Depth and breadth of GII HuNoV genome coverage in the enrichment optimisation investigation (NE=No enrichment, PA=Poly-A tail selection, P1=Single PGM-MB enrichment, P10=Pooling of ten PGM-MB reactions).

Sample (Treatment)	Library PCR cycles	Library Ct value	DNA (ng/µL)	Reads mapped to genome (%)	Percent genome (%)	Coverage	
						Mean	Standard deviation
15-13 (NE)	27	16.57	1.07	25.7	99.98	3599.58	1461.43
15-13 (PA)	12	14.56	5.46	46.38	100	4009.15	1653.85
15-13 (P1)	27	17.4	1.62	14.02	100	2302.74	1091.33
15-13 (P10)	12	10.13	7.21	61.94	100	5224.45	1769.75

3.3.7 Deviations in HuNoV consensus genome coverage between treatments

In this pilot investigation, enrichment methods were optimised using a single sample of high HuNoV titre. The coverage plots of sample A were similar in shape after alignment to the GII HuNoV consensus, regardless of the treatment used (Figure 3.9).

To understand the influence of sequence composition on HuNoV library preparation, a 50-base sliding window of GC content was plotted against mapped read coverage to the HuNoV consensus genome for all sample treatments (Figure 3.10). In the graph, coverage decreased gradually from 30-66% GC content, but there were two exceptions at 28 and 68% where coverage dropped off sharply (Figure 3.10).

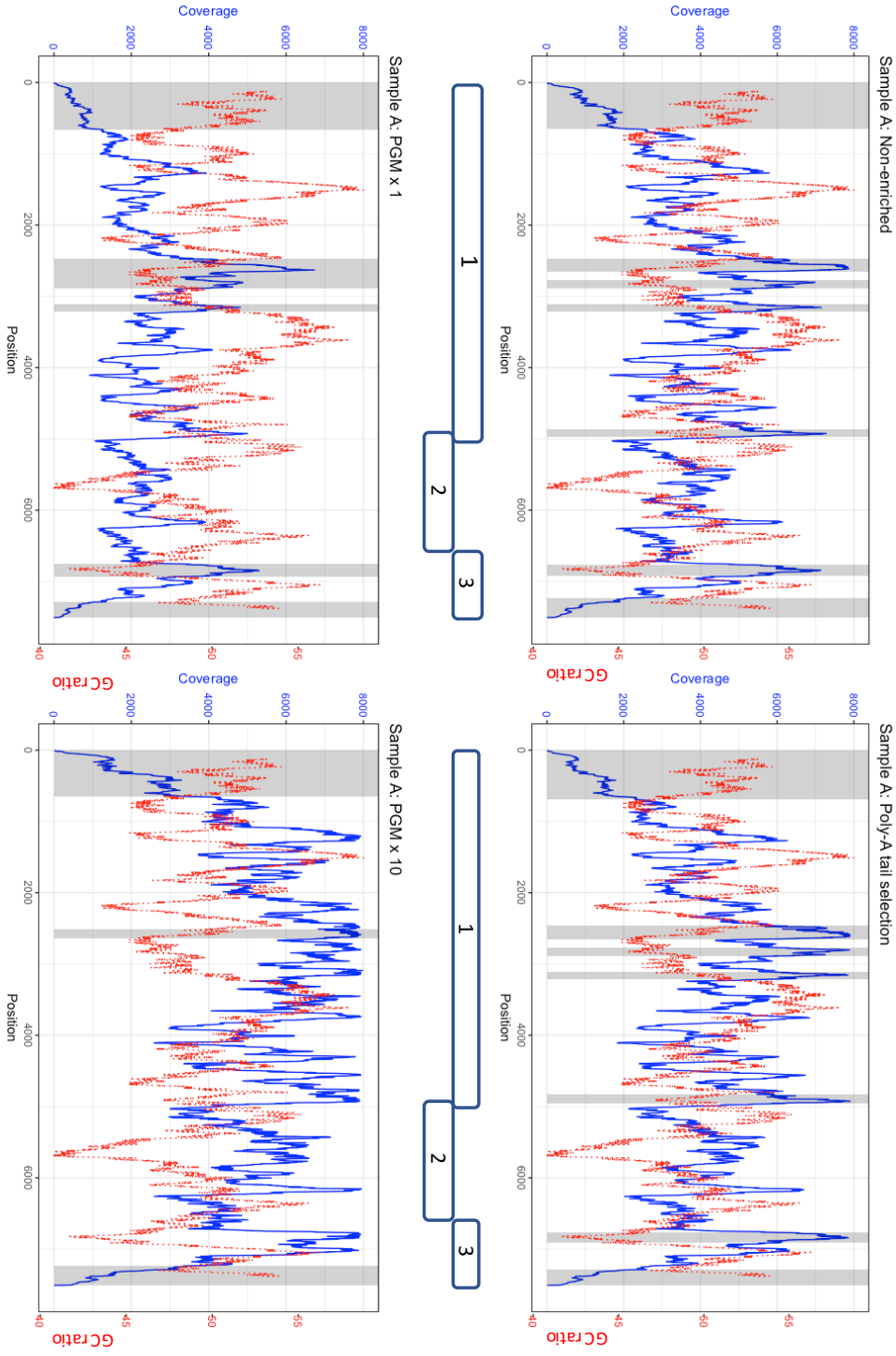


Figure 3.9 HiSeq read coverage plots of the HuNoV consensus genome for sample A (Red line: 250 base sliding window of GC content, blue line: Coverage of base position and shaded regions are above or below one SD of the mean coverage).

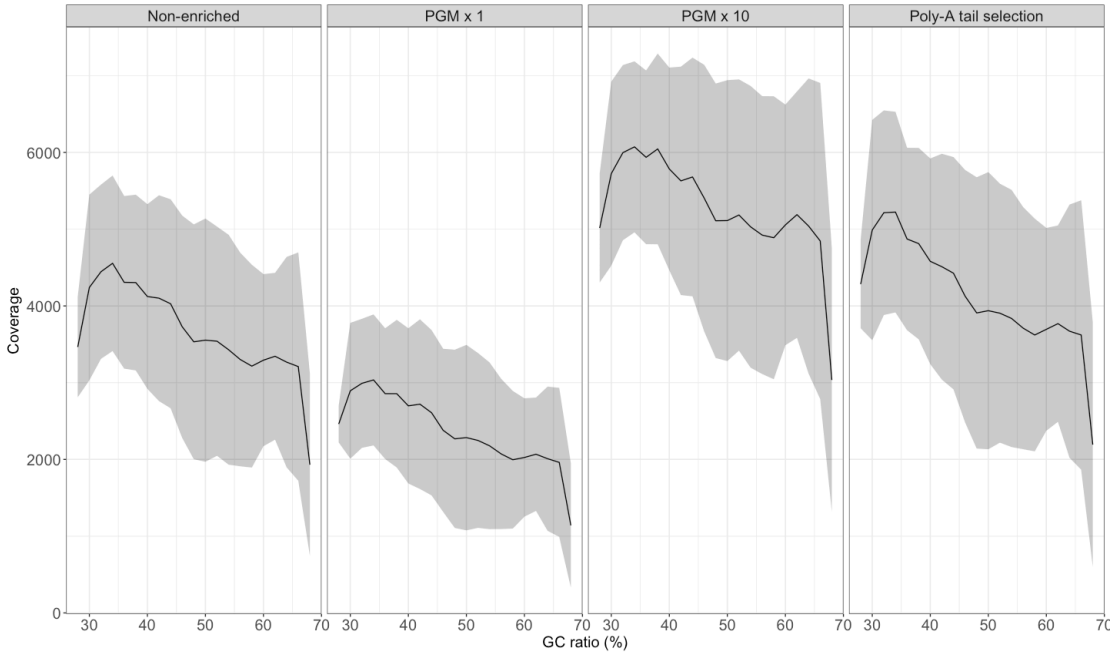


Figure 3.10 HiSeq read coverage plotted against GC ratio (Black line:50 bp sliding window, grey areas: Standard deviation).

Different regions in the genome (≥ 100 bases) were examined more closely for abnormalities in coverage (over 1 SD). The read coverage dropped for all samples at the start and end of the HuNoV consensus genome (Figure 3.9 and Table 3.11). In ORF 1, there were four internal regions of the genome with read coverage abnormally higher than the mean (over 1 SD), however only region B followed this trend in all sample treatments (Table 3.11). In contrast, none of the regions in ORF2 deviated above or below the mean, and region F in ORF3 showed a deviation above the mean in three quarters of the sample treatments. The majority of regions that had an abnormally high coverage had a lower average GC ratio in comparison to the mean content of the whole consensus genome (50.1 %), which was used to align the reads of the different sample treatments (Table 3.11).

Table 3.11 Deviations in coverage across the GII HuNoV genome (Ne=Non-enrichment, PA=Poly-A tail selection, P1= PGM × 1, P10= PGM × 10).

Region	Region position range in different libraries	ORF	GC content (%)	Above/Below standard deviation	Sample
A	1-694	1	51.6	Below	NE, PA, P1, P10
B	2458-2727		48.8	Above	NE, PA, P1, P10
C	2766-2893		52.1	Above	NE, PA, P1
D	3110-3216		48.9	Above	NE, PA, P1
E	4830-4969		48.1	Above	NE, PA
F	6756-6931	3	42.0	Above	NE, PA, P1
G	7233-7509		53.3	Below	NE, PA, P1, P10

3.3.8 Metagenomics of PGM enriched samples

Molecular quantification proved, in agreement with the literature, that GII HuNoV could bind PGM on the surface of MBs. However, binding to carbohydrate molecules and PGM is not HuNoV specific, and other enteric pathogens such as rotavirus and *Campylobacter* species can also bind to this class of molecules (Hu et al. 2012; Ruiz-Palacios et al. 2003). Further investigations were performed, aimed at examining the metagenome of the clinical samples subjected to PGM-MB capture. The samples used for this included clinical stool samples, positive for GII HuNoV, collected from three patients on different days. The collection of samples underwent PGM-MB capture and MPS in parallel with a separate GII NoV positive clinical sample which received no PGM-MB treatment. A similar number of reads were generated in all but one of the samples, the exception, sample 11-3, had a

read count three magnitudes lower and could be a result of failure during library preparation or sequencing (data not shown).

In the remaining samples, the largest number of the reads mapped to viruses, of which a greater proportion were non-bacterial viruses (Figure 3.11). In most of the samples collected PGM-MB capture was effective, and the greatest proportion of the non-bacterial virus group mapped to each respective HuNoV sample consensus sequence (Figure 3.12). However, 3 out of 10 PGM-MB treated samples (4-3, 13-4 and 13-5), which did not appear to fail library preparation or sequencing had below 2% of the reads mapped to their GII HuNoV consensus sequence (Figure 3.11). The PGM untreated stool contained very little HuNoV, and therefore corresponded closely to another sample which was considered to have failed library preparation (Figures 3.11 and 3.12).

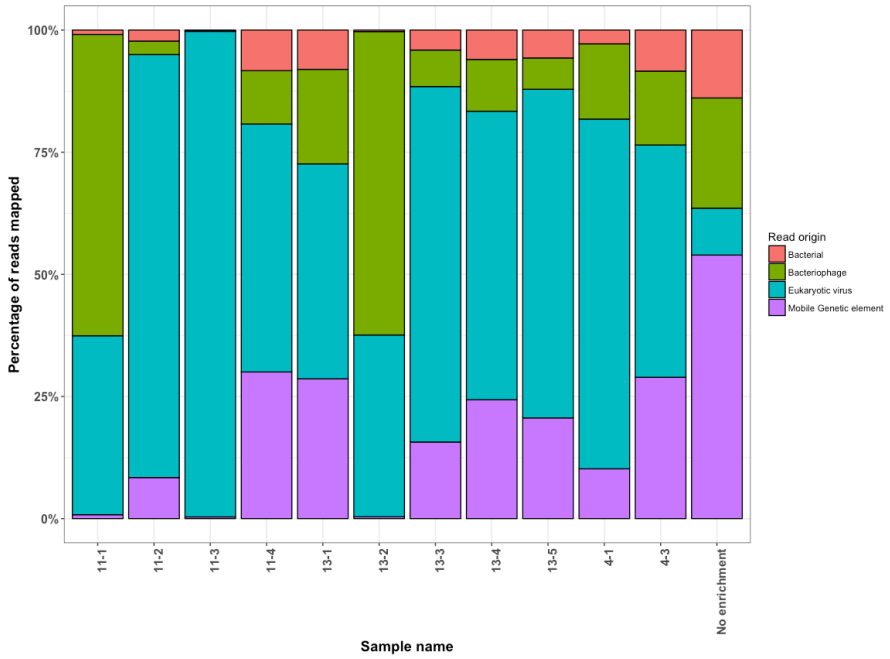


Figure 3.11 MPS reads (Illumina) mapped and grouped by a RefSeq database, from longitudinal HuNoV clinical samples positive for GII HuNoV, with or without PGM capture.

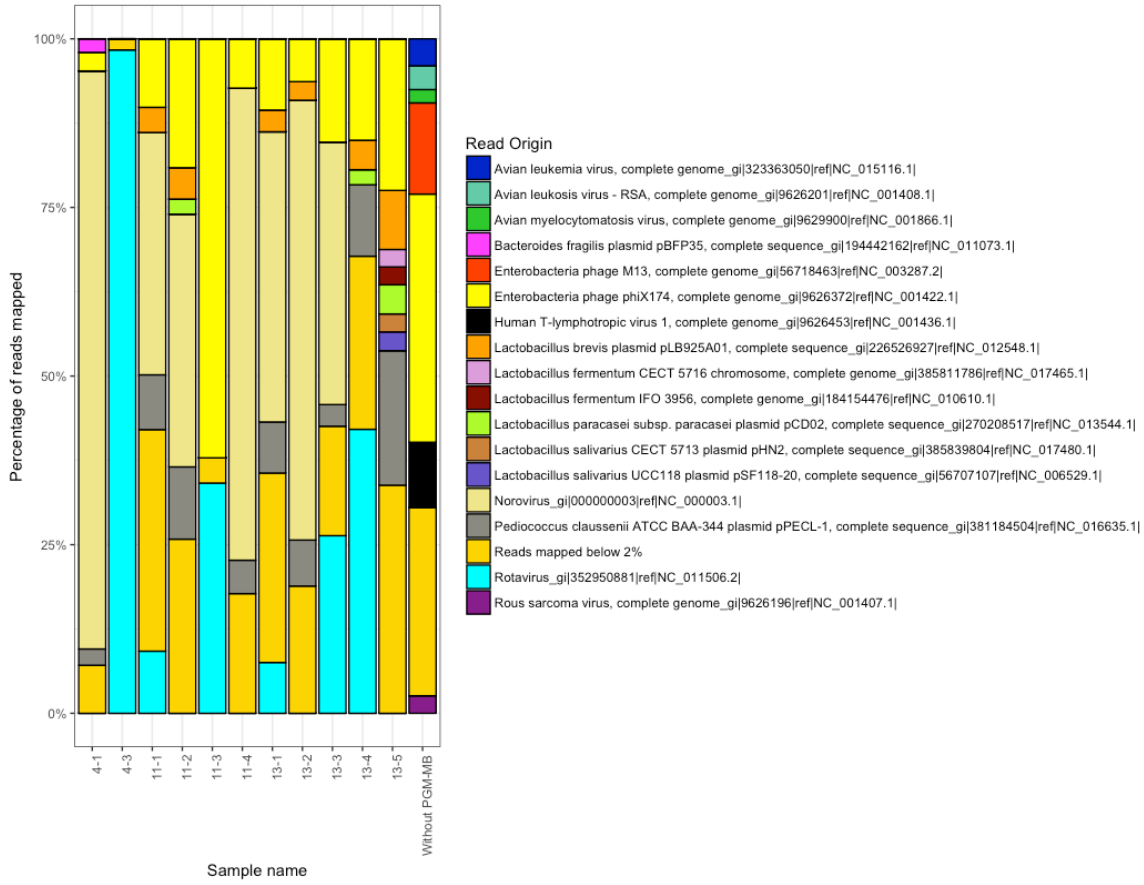


Figure 3.12 MPS reads (Illumina) mapped and typed by a RefSeq database, from longitudinal HuNoV clinical samples positive for GII HuNoV, with or without PGM capture.

The effects of PGM-MB enrichment or additional PCR cycling upon the metagenomics of GII positive stool samples was investigated. The high (15-13) and low (17) titre clinical samples were compared directly, under the influence of PGM-MB enrichment and additional PCR cycling (Figure 3.13). The greatest enrichment effect was observed in the high titre sample with PGM-MB capture, whilst a combination of PCR and PGM-MB capture led to less reads in comparison (Figure 3.13). Interestingly, a decrease in the proportion of HuNoV reads was detected if additional PCR cycling was performed in contrast to no treatment (Figure 3.13). In the weak titre sample, both PGM-MB capture and additional PCR

cycling were unable to enrich the proportion of HuNoV reads above 2% (Figure 3.13).

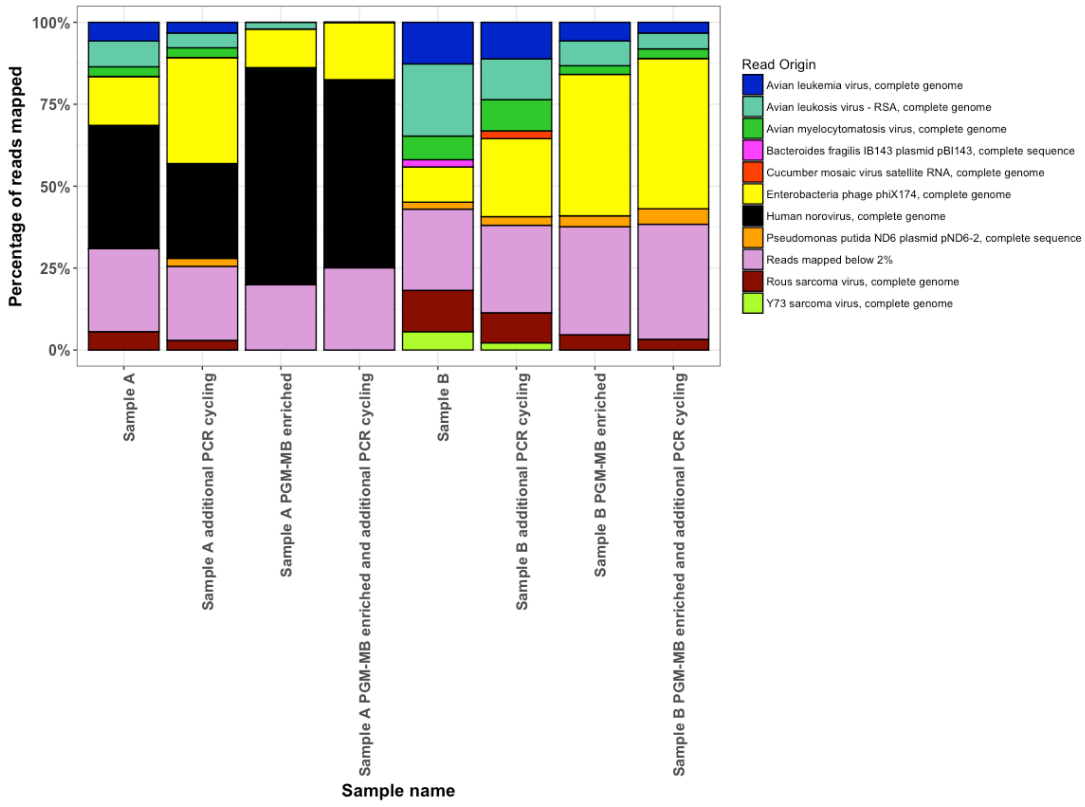


Figure 3.13 MPS reads (Illumina) mapped and grouped by a RefSeq database, from a high (Sample A=15-13) and low (Sample B=17) titre HuNoV clinical sample positive for GII HuNoV, with or without a combination of PGM capture and additional PCR cycling.

3.3.9 The effects of enrichment and PCR on variant bias

In the initial experiments to investigate the effects of poly-A tail enrichment and PGM-MB capture on GII HuNoV, a subset of clinical samples required additional non-specific amplification for even nucleic acid pooling prior to MPS. Investigations thus far have shown several PGM-MB captures were more effective than poly-A tail enrichment at increasing the availability of GII HuNoV reads (Table 3.10), and

therefore further experiments were performed to understand the effects of capture and non-specific amplification on library quality and minority variant frequency.

In the high titre sample (15-13), each treatment recovered a near complete consensus genome (Table 3.12). However, the greatest mean coverage was obtained by PGM-MB capture, which was followed by capture and additional non-specific amplification (Table 3.12 and Figure 3.14). If the number of library preparation cycles doubled, with or without PGM-MB capture, the total DNA concentration decreased alongside the mean GII HuNoV coverage (Table 3.12).

In the lower titre sample (17), PGM-MB capture did not increase the fraction of the consensus genome recovered in comparison to no treatment (Table 3.12). Enrichment by additional non-specific amplification increased genome recovery by 16.2%, and a combination of capture and additional cycling recovered a near complete consensus genome (Table 3.12 and Figure 3.15). The recovery of a near complete genome coincided with the greatest mean coverage for the low titre sample (Table 3.12). However, in the low titre sample with PGM-MB capture, and the high titre sample with no treatment, a high library DNA concentration did not correlate with HuNoV mean coverage (Table 3.12).

Table 3.12 The effects of non-specific amplification and PGM-MB capture on the total DNA concentration after the library preparation stage, the GII HuNoV read coverage (IQR=Interquartile range) and differences in read mapping.

Sample	Treatment	Library preparation PCR cycles	Percentage of HuNoV genome (%)	DNA concentration (ng/µL)	Coverage		Percentage of total reads mapping to HuNoV consensus (%)	Fold different realtive to no treatment
					Median	IQR		
15-13 (High titre)	None	10	99	7.24	30	31	0.14	
	PGM-MB capture		99.1	7.35	82	72	0.4	2.9
	Additional PCR cycling	20	97.9	5.04	20	16	0.07	0.5
	PGM-MB capture and additional PCR cycling		99	6.85	51	57	0.18	1.3
17 (Low titre)	None	10	37.9	2.21	2	2	< 0.01	
	PGM-MB capture		27.9	7.11	2	1		0.5
	Additional PCR cycling	20	54.1	4.05	2	2		0.3
	PGM-MB capture and additional PCR cycling		99.2	5.32	6	7		0.8

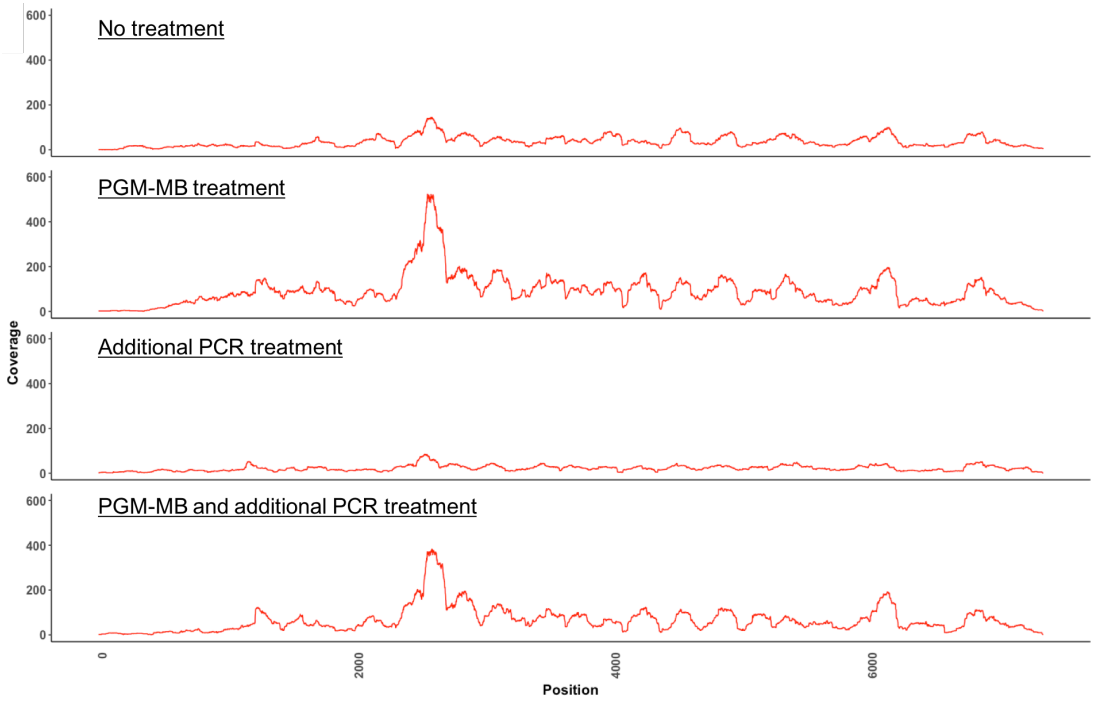


Figure 3.14 HiSeq read coverage plots of a high titre GII HuNoV clinical sample after PGM-MB and/or non-specific PCR treatment.

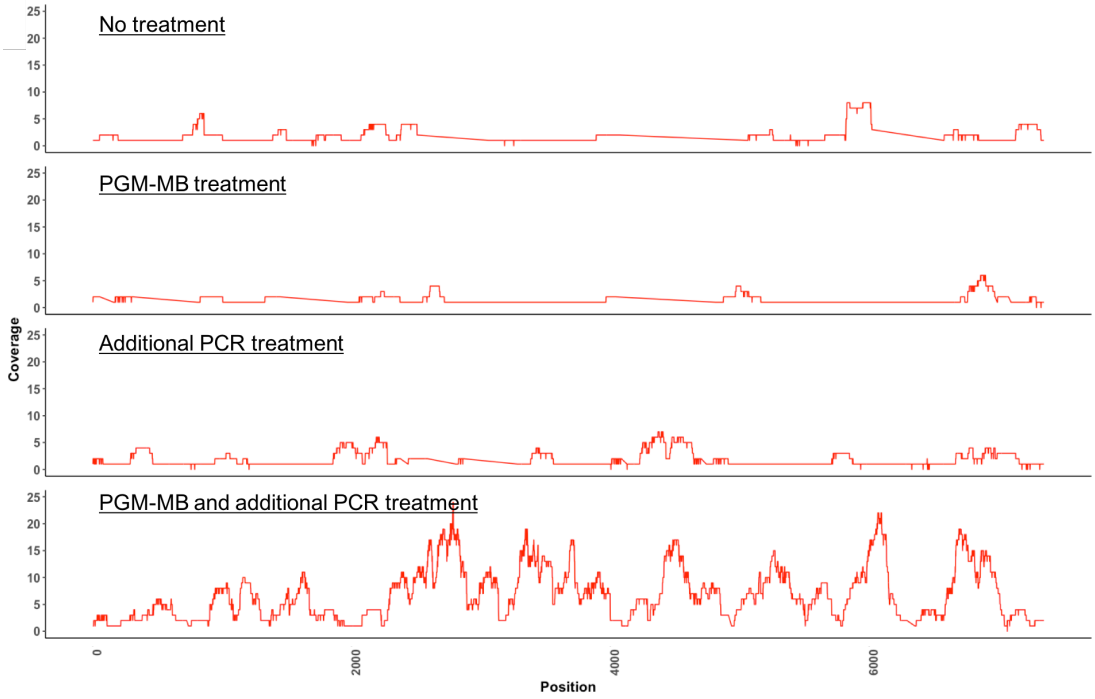


Figure 3.15 HiSeq read coverage plots of a low titre GII HuNoV clinical sample after PGM-MB and/or non-specific PCR treatment.

The fold increase in percentage of total reads mapped for the higher titre clinical sample (15-13) reflected the pattern observed in mean coverage (Table 3.12). If PGM-MB was applied the proportion of reads mapping to the consensus increased 3-fold, whilst additional PCR cycles with or without capture decreased the fold difference to 1.3 or 0.5, respectively. On the other hand, PGM-MB capture, PCR or both led to a decrease in the relative proportion of HuNoV reads when compared to the untreated low titre sample (Table 3.12).

The high and low titre sample, and each of their treatments, were sequenced on the same Illumina flow cell and given error rates are constant between each sample it should be possible to assess the effects of capture and non-specific amplification on the frequency of minority variant base calls.

To assess whether any skewing existed in the frequency of minority variants, a subset of positions in the sequence data of the high titre sample were examined. A position was considered for comparison if, firstly, a minority base was present in more than one sample treatment and, secondly, if the base had been called twice or more (Figure 3.16). Most positions increased in variant frequency in the presence of PGM-MB capture treatment, and if it was combined with extra PCR cycling (Figure 3.16). Therefore, further analysis was performed to measure whether any skew existed between each untreated sample (15-13 and 17-1), and each treatment (Figure 3.17).

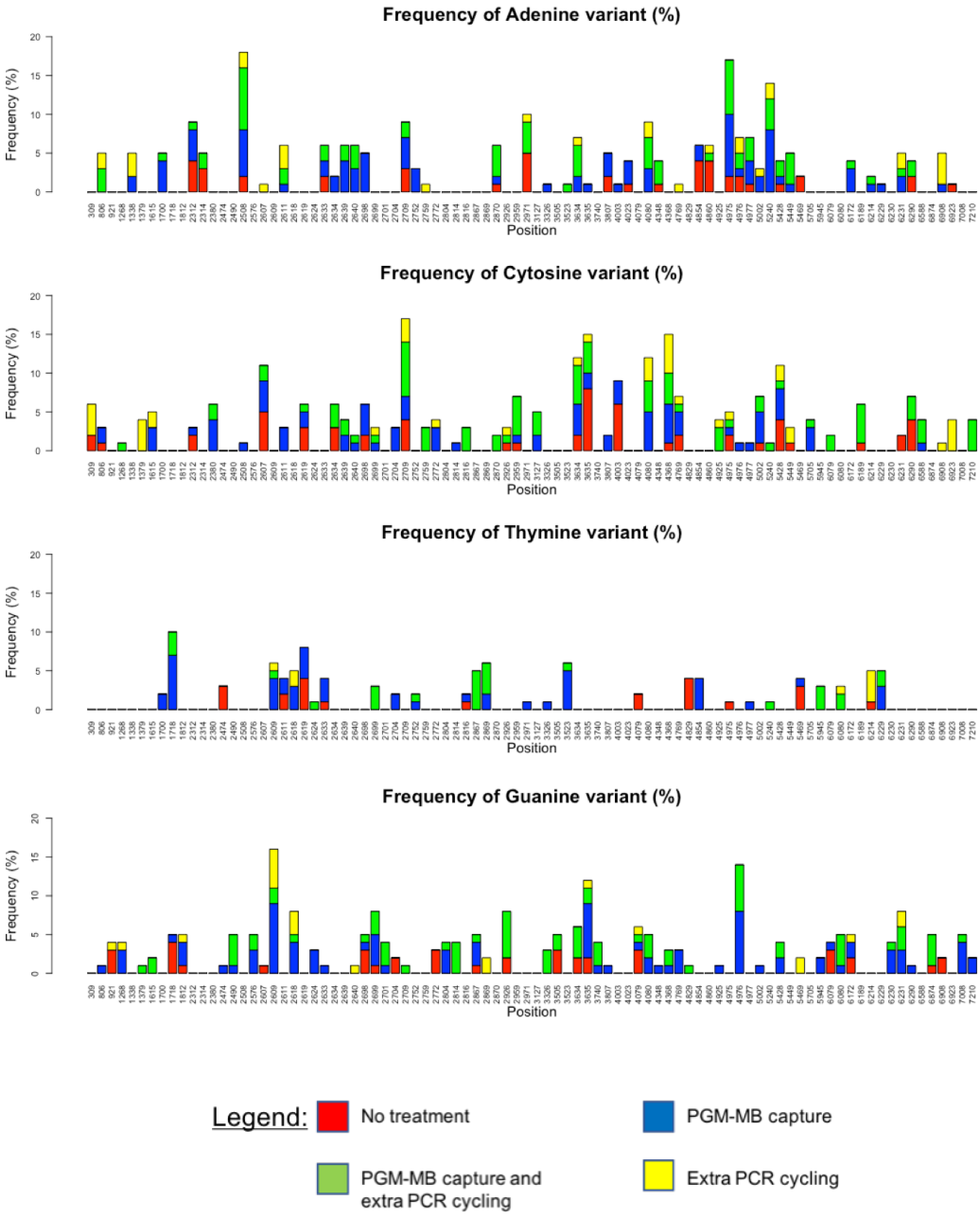


Figure 3.16 The skew of minority variants at a subset of positions across the consensus sequence obtained from the high titre sample.

If the high titre sample underwent PGM-MB capture or additional PCR cycling combined with capture the frequency of minority variants at each respective position appeared to be even (Figure 3.17). But if extra PCR cycling was applied solely, the frequency of variants became skewed toward the untreated sample, implying additional non-specific amplification decreases the ability to detect of

minority variants at each position (Figure 3.17). In the lower titre sample a limited number of variants were present, presumably due to the low coverage and the absence of a complete genome in most samples (Table 3.12 and Figure 3.17).

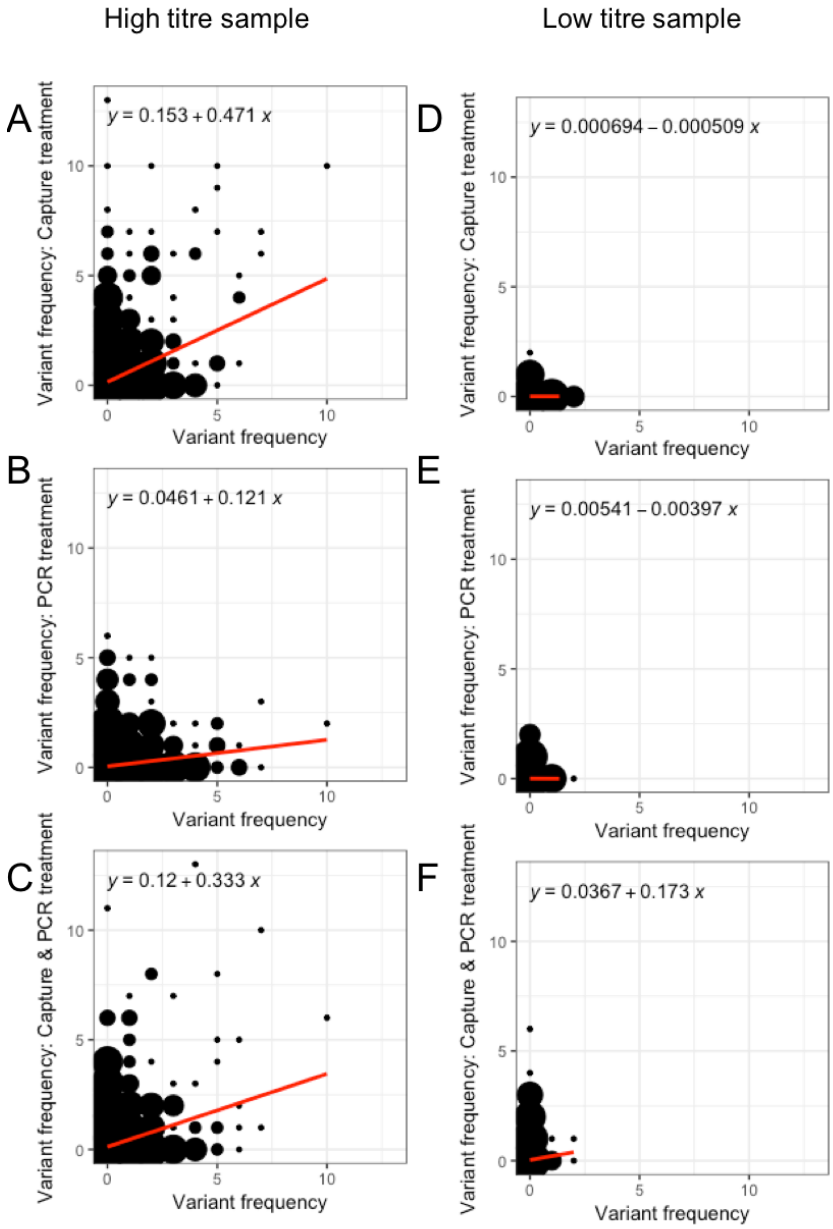


Figure 3.17 The frequency of variants called after treatment (y-axis) against the frequency of variants present in the non-enriched sample (x-axis). A and D = PGM-MB capture, B and E = Additional PCR amplification, C and F = PGM-MB capture and additional PCR amplification.

3.4 Discussion

3.4.1 Optimisation of PGM GII HuNoV capture

In a clinical stool sample, HuNoV represents a very small proportion of the total nucleic acid material (Nakamura et al. 2009). To examine the whole genomes of a HuNoV population in this environment it is necessary to increase the virus representation or remove the 'contaminants'. These investigations demonstrate the validity of a HuNoV enrichment method that relies on capturing viral particles, and is based on the well described interactions between the virus capsid and HBGAs. One advantage of this method is that PGM can capture HuNoV strains regardless of genogroup or genotype specificity, and therefore provides a universally applicable method (Tian et al. 2010).

3.4.2 The PGM-MB metagenome

HuNoV has a proven affinity for the HBGA structures on the villus surface, and as amply described in the literature, the expression of different HBGA influences host susceptibility to the virus (Huang et al. 2005; Marionneau et al. 2002). PGM is a mixture of over 30 diverse glycosylated protein structures; but some of these are analogous to the HBGAs in humans, and this has been shown by reactivity with a mAb known to recognise a HBGA type A conformational structure (Karlsson et al. 1997). Subsequently, the ability of the purified PGM (Glenister et al. 1988) to compete with HBGA for HuNoV VLP binding, in a dose dependent manner, was proven by Tian and colleagues (Tian et al. 2005). The HBGA structures present in commercial PGM type III were further characterised by Lindesmith and colleagues, and shown to contain H, A and Lewis Y antigens, which further explained the broad

binding affinity of this mixture to HuNoVs (Huang et al. 2005; Lindesmith et al. 2012b). This mixture of proteins can be coated onto MBs and used to concentrate HuNoV (Tian et al. 2008).

PGM-MB capture experiments of GII HuNoV were optimised, under different environmental parameters, on two serial dilutions of a clinical sample, corresponding to strong and weak viral loads. At a low pH, there was a trend of increased PGM-MB capture efficiency, which was significant when applied to a sample with low viral load ($p = < 0.05$). When the duration of PGM-MB exposure to the virus was increased, there was no change in the yield, which indicates the GII HuNoV capture occurs shortly after being in contact with the HBGA-like molecules, and extension of incubation time does not have an impact on capture efficiency. Both findings agree with previous work (Tian et al. 2010) in which PGM-MB capture of GII.4 HuNoV was found to be rapid, and a significant increase in virus binding was observed from a sample of weak viral load at pH 3.6. Tian and colleagues postulated that this increase in affinity is due to the overall charge of PGM (negative) and the GII.4 HuNoV VLP (positive) at low pH (Tian et al. 2010). On the other hand, if the temperature was altered from 4°C to 37°C, no significant change in virus yield occurred at high viral load, but for the diluted sample, incubation at 4°C significantly decreased the yield of GII HuNoV ($p = < 0.05$). A reduction in HuNoV yield at lower temperatures could be a result of a reduced reaction rate between the virus surface receptor and PGM.

There was some variability in the results obtained with PGM-MBs capture applied to different clinical samples, which did not necessarily correlate with differences in

viral load. Stool samples comprise a complex matrix; in addition to the virus of interest, which may be found clumped, free or in a cell associated form, it will comprise unabsorbed food, luminal secretions, the host microbiota and the host virome. A loss of capture was observed in sample 13-4 and 13-5 (Figure 3.12) and could have been due to progression of the infection to the convalescent stages. In the convalescent stages of infection, various factors in addition to the decreasing viral load are likely to hamper the HuNoV capture efficacy. Secretions into the lumen will comprise of host mucins, secretory antibodies and cell debris from regular epithelial cell shedding (Bullen et al. 2006). The cell debris and mucins will be partially composed of glycosylated proteins, or HBGAs with affinity for HuNoV particles present in the stool. In addition, if the host has seroconverted, the secreted antibodies will mask or neutralise the virus surface receptors, making them unavailable to the mucin bait.

HBGA-like structures are found on the surface of gram-negative bacteria, and therefore the microbiota of some individuals may bind and capture HuNoV to interfere with a mucin based capture method. Bacterial HBGA-like structures were first described in 1961, and provided an explanation for non-self-blood group Ab activity (Springer et al. 1961). These HBGA-like structures were more recently found on *Escherichia coli, Escherichia fergusonii*, *Enterobacter cloacae* and *Shigella flexneri* by screening bacterial isolates or stool samples for reactivity to anti-blood group specific Abs (Li et al. 2015; Miura et al. 2013). Furthermore, it was possible to elucidate a direct interaction between the extracellular polymeric substance of *E.cloacae* and HuNoV VLPs (Miura et al. 2013). Subsequently, the rescue of HuNoV

GII.4 Sydney infection of BJAB B cells after incubation with *E.cloacae* is in agreement with these bacterial HBGA-like molecule-HuNoV interactions (Jones et al. 2014).

Despite the efficacy of mucin, and HBGA-like molecules within, for capturing HuNoV, this is not an exclusive interaction, and other viruses and various bacteria are also known to bind HBGA. The data presented here demonstrated PGM-MB capture provided greater proportions of various bacteriophages, mobile genetic elements and other eukaryotic viruses, but effectively reduced the detection of bacterial DNA. The proportion of reads that were mapped to HuNoV were significantly higher in the stool samples subjected to the mucin capture method than in those sequenced without prior mucin capture. The impact of this enrichment is especially significant when dealing with samples of low viral load.

Therefore, due to PGM-MB interactions not being strictly confined to HuNoV and the above factors, this method, whilst providing a reasonable enrichment of HuNoV from clinical samples, is subject to a degree of interference and variability that cannot be entirely controlled. However, this technique can increase the representation of HuNoV nucleic acid over other 'contaminant' sequences, prior to MPS, even if it does not exclude other organisms entirely or serve to concentrate the virus effectively, unless used in parallel (PGM-MB pooling).

The method described here offers numerous advantages over the use of PCR before library preparation and MPS. Firstly, PGM-MBs should exclude most components in the clinical samples unable to bind to the receptor. Secondly, clinical samples with very low HuNoV titres can be enriched, by pooling, without the use of PCR which

could introduce a bias. This means that the efficacy of this method should not be impacted by the emergence of new HuNoV genotypes or variants, and will also capture and sequence mixtures of HuNoVs in a sample.

3.4.3 Limitations of enrichment protocols

Several effective enrichment protocols for HuNoV, prior to MPS, have been published in the literature. However, protocols can be limited by an inability to discriminate between viral and non-viral nucleic acids (poly-A tail enrichment), or by targeting a single virus type (RNA bait enrichment). If a nucleic acid extraction is performed on stool samples, host mRNAs can compete with the viral genomic RNA during poly-A tail enrichment, and reduce the availability of HuNoV sequence data downstream. On the other hand, RNA baits are short oligonucleotide fragments designed prior to MPS, to enrich for a virus after extraction. However, if the sample contained a mixture of HuNoVs or a virus recombinant, the whole virus population will not be selected by enrichment and be unavailable for downstream analysis (Brown et al. 2016). RNA bait enrichment may particularly underrepresent minority species.

3.4.4 The effects of library preparation and NGS on GII HuNoV read mapping

In this pilot investigation, a deviation away from the mean coverage could be seen in several regions (> 100 bases), and the sizes ranged from ~106 to 694 bases. The deviation in coverage was below average at the start and end of the genome for all library preparations (Figure 3.9). There are two variables that are likely contributors to this observation. Firstly, the Illumina library preparation process is limited by an inability to recover short dsDNA fragments that would be generated at the start and

end of the genome, and this has been described previously for HuNoV (Batty et al. 2013). Secondly, the 5' end, ORF1/2 junction and the 3' end have been predicted to form secondary structures, which could induce a steric hindrance prior to cDNA synthesis or PCR amplification (Simmonds et al. 2008). The ORF1/2 junction is not completely highlighted in this analysis (Figure 3.9), but a sudden decline in coverage did occur from positions 4944-4971 (> 1 SD above the mean) to positions 5020-5023 (>1 SD below the mean) in all library preparations. Simmonds and colleagues conservatively predicted the presence of secondary structures at three regions across the genomic sense strand of different viruses within the *Caliciviridae*, which included GII.4, by multiple bioinformatics algorithms (Simmonds et al. 2008). The algorithms detected a drop in synonymous site variability and a change in the minimum free energy difference, when compared against randomly scrambled subsequences, at the 3' and 5' region of each genome, as well as upstream of the ORF1/2 junction (Simmonds et al. 2008). Moreover, the algorithms agreed with secondary structure prediction software at the ORF1/2 junction and 5' end of the GII.4 genome (Simmonds et al. 2008).

In the remaining regions, there was an overrepresentation of reads (>1 SD above the mean) in the first (4 regions) and third (1 region) ORFs (Table 3.11). The unusually high coverage in these regions could be attributed to an absence of predicted secondary structures or a bias in the library preparation process, which can be induced by extremes of GC content, and was lower than average in four of five regions (Simmonds et al. 2008). During library preparation, the source of bias could be fragmentation, if the reaction is not sequence independent, at a primer

annealing stage, if the affinity is not random, or because of stochasticity in the PCR step. The fragmentation step in the ScriptSeq (Illumina) kit is chemical, and in contrast to enzymatic methods likely to be sequence independent (Hansen et al. 2010). The subsequent RT step is initiated by random hexamers which, in transcriptomics, have been shown to bind selectively to regions adjacent to specific nucleotide profiles of 13 nt (Hansen et al. 2010). Moreover, those RNA fragments converted to cDNA have undergone PCR enrichment, and this step has been identified as a significant source of base-composition bias (Aird et al. 2011). Aird and colleagues found base composition did influence coverage, and regions with a higher than average GC content were amplified less efficiently during the library preparation stage (Figure 3.10). After PCR enrichment, the relative abundance of amplicons with a GC content of, 12% or below, or, greater than 65%, decreased to ~1 tenth or ~1 hundredth of the expected abundance, respectively (Aird et al. 2011). The bias was largely attributed to differences between thermocyclers and the ramp rate, and by slowing the ramp rate, to allow for more denaturation time, it was possible to recover the relative abundance of templates with a high GC content (Aird et al. 2011). Therefore, if the GC ratio is high, the dsDNA does not separate efficiently and the hexamer cannot bind to amplify the template, but if the ratio is too low there is inefficient annealing and extension. The peak of average base coverage at different GC contents was 34% GC, and this could reflect the lesser number of hydrogen bonds between bases needed to be denatured in an AT rich template (Figure 3.10). Other investigations have found a correlation of increasing read coverage with an increase in the GC ratio of reads. However, this

finding was generated with the Solexa instrument (Illumina), which produces shorter reads (Dohm et al. 2008; Hillier et al. 2008).

3.4.5 The effects of random PCR amplification bias on the frequency of variants

A PGM-MB capture method is effective in a high titre stool sample, however can be limited if the viral load is low or inaccessible (Figure 3.12). If extra non-specific PCR cycling or PGM-MB capture were applied to the low titre sample the fraction of HuNoV genome recovered was 54.1% or 27.9%, respectively. In contrast, when the low titre sample underwent PGM-MB capture and non-specific amplification the virus genome recovery increased to 99.2%, and implies synergy between the two different methodologies. On the other hand, capture recovered near complete virus genomes in a stool sample of high viral load. Moreover, additional non-specific amplification constrained the sequence diversity available in a sample of high HuNoV viral load in comparison to PGM-MB capture. This is likely due to PCR selection and drift leading to the high efficiency exponential amplification of a subset of templates in relation to others (Wagner et al. 1994). The effects of amplification skew on MPS has been observed previously by the assignment of unique primer identifiers to a known number of HIV cDNA templates through degenerate priming, prior to 454 (Roche) MPS (Jabara et al. 2011).

3.4.6 Limitations of metagenomics analysis

A limitation of this metagenomics analysis would be the lack of positive and negative controls. In a shotgun metagenomics experiment negative controls are "blanks" during sampling, nucleic acid extraction and library preparation (Kim et al.

2017). Therefore, to eliminate any contaminating sequences during a HuNoV PGM-MB capture, it would be necessary to perform MPS on PGM-MBs without a clinical sample, on the eluate from the extraction procedure and on a library preparation without template. Similarly a positive control could have been included in all experiments prior to nucleic acid extraction and library preparation to ensure efficiency at each stage (Kim et al. 2017). A lack of negative controls has been a significant issue in metagenomics studies of low biomass samples, and misleading results have been directed by the nucleic acid extraction kit chosen rather than the experimental variables (Salter et al. 2014; Turner et al. 2012). However, contaminant nucleic acid sequences originating from reagents can also be detected in high biomass samples, which cannot be arbitrarily discarded without a negative control (Glassing et al. 2016; Laurence et al. 2014). To avoid misleading results due to contamination negative controls can be introduced, to remove contaminant nucleic acids during bioinformatics analysis, or laboratory reagents can be treated with ultraviolet (UV) light, enzymes or ethidium monoazide (EMA) to minimise the contaminant nucleic acids carried forward to MPS (Corless et al. 2000; Glassing et al. 2016). The stool samples used in these investigations would be considered high biomass, and less likely to be influenced by contamination. However, further experiments should investigate whether contamination is present prior to capture, which could have an influence on receptor saturation on the surface of the PGM-MBs.

Another limitation to consider is the potential for “spreading-of-signals” or “adapter leakage” directly prior to MPS of the pooled samples. “Spreading-of-

signals" has been described on the HiSeq platform and causes 5-10% of reads to be assigned to the incorrect sample index (Sinha et al. 2017). This is due to low levels of free index primers, which are carried forward by AMPure XP beads (Beckman Coulter) and complementary to other indices in regions other than the barcode. These primers allow DNA polymerase to extend the library at the 3' end to generate an identical library strand with a different index (Sinha et al. 2017). The metagenomics investigations performed were on the same flow cell, and therefore, 5-10% of reads in a sample could be assigned to the incorrect patient sample or treatment.

4 HuNoV diversity in longitudinal samples over the course of infection

4.1 Introduction

4.1.1 Evolution of RNA viruses

The persistence of a virus in the population is dictated by its ability to infect a host and generate transmissible progeny; which can induce further infections by intra- or inter-host transmission. Therefore, experimentally, the infectious progeny titre of an RNA virus is measured as a proxy for viral fitness (Domingo et al. 1997). An RNA virus replicating in a host exists as a population of randomly mutating swarms, or quasispecies, and if a selective pressure is present swarms with higher fitness will become more prevalent and lead to evolution of the population. However, more recently, investigations have implied that quasispecies dynamics are more complex, and that interactions between viruses in the population have a role in the maintenance of high or low fitness variants, by complementation or interference (Andino and Domingo 2015).

Complementation (positive interaction) or interference (negative interaction) occurs when a subpopulation of the same quasispecies interacts with another to promote or inhibit replication (Andino and Domingo 2015). The existence of complementation in virus populations has been shown with a Poliovirus mutant population (G64S), which replicates at high fidelity and therefore is limited in the diversity it can generate after replication when compared to the wild-type (Vignuzzi et al. 2006). Vignuzzi and colleagues proved that limited diversity in the G64S

population precluded neurotropism in mice, however the presence of a diverse wild-type or G64S population recovered the neurotropic phenotype (Vignuzzi et al. 2006). On the other hand, subpopulations that interfere with replication, or defectors, have been demonstrated by electroporation of Foot-and-mouth disease virus (FMDV) wild-type RNA with capsid and polymerase mutants, which subsequently led to a relative decrease in the wild-type titre (Perales et al. 2007).

4.1.2 MPS as a tool to infer virus population dynamics in a host

The availability of MPS platforms has facilitated broad, in-depth and high-throughput analysis of virus swarms, through the alignment of many highly repeated and overlapping short sequence fragments encompassing the whole viral genome. This contrasts with previous sequencing technologies (e.g Sanger sequencing) that relied on sequential generation of relatively short regions of consensus sequence, that are limited in their detection of minority variants (Palmer et al. 2005). MPS is a highly sensitive tool which can be used to infer relative viral load, virus interactions with the host or other pathogens, genetic linkage, virus mutation rate or regions of the virus genome under selective pressure (Thys et al. 2015; Töpfer et al. 2013; Zagordi et al. 2011).

4.1.3 The impact of MPS on public health and disease control

Several different types of MPS platform have demonstrated the diversity present in RNA virus swarms, and how the frequency of minority variants can change over time. MPS of RNA viruses has been applied to study population dynamics to inform on: adaption to cross-species barrier transmission, efficacy of antiviral drug regimens in clinic, hypervariable positions in the virus genome and the genetic

stability of live vaccine formulations (Hasing et al. 2016; Jonges et al. 2013; Majid et al. 2015).

4.1.3.1 Detection of a minority variant adaption to cross-species transmission

MPS has been used to successfully identify markers for human host adaption present in minority variants of Influenza A virus. Influenza A viruses are thought to exist in wild aquatic birds as reservoir hosts, and mutations can occur that allow adaption to mammalian host infection (Yoon et al. 2014). Avian influenza viruses that have adapted to infect humans replicate more efficiently in a lower temperature environment, the upper respiratory tract (~33^{o}C), when compared to the gastrointestinal tract of wild aquatic birds (42^{o}C). This has been demonstrated *in vitro* by the transfection of plasmids expressing Influenza A virus (IAV) subunits into mammalian cells alongside a chloramphenicol acetyl transferase (CAT) expressing plasmid with a virus-like promoter (Massin et al. 2001). During avian influenza virus infection of mammalian cells, transcription and replication of CAT was noted to be recovered at 33^{o}C by a compensatory amino acid (AA) change in the PB2 subunit (E627K) (Massin et al. 2001). Jonges and colleagues, using MPS, demonstrated the presence of the E627K mutation as the majority variant in a mixed H7N7 virus population after zoonotic transmission to a human (Jonges et al. 2014). The majority E627K variant was detected in bronchoalveolar lavage (75%) and sputum samples (93%); reaching 100% frequency after five days. However, the mutant could not be detected in poultry prior to the transmission event (Jonges et al. 2014).

4.1.3.2 Identification of minority variants as an indicator of poor antiviral drug efficacy

MPS platforms could be an important clinical tool in monitoring how combination therapy controls chronic hepatitis C virus (HCV) or human immunodeficiency virus (HIV) infection, and reducing the likelihood of antiviral resistant minority variants from emerging. The sensitivity of MPS is advantageous in chronic viral illness as it is possible to detect emergent resistance mutations early, which could be selected for in chronic infection and lead to treatment failure. However, the relationship between the presence of minority variants with antiviral resistance and the patient outcome on clinical therapy is controversial. HCV populations in individuals treated with interferon and ribavirin consisted of minority resistance mutations to direct acting antivirals (polymerase & protease inhibitors) without exposure, implying a resistant genotype occurs at random (Nasu et al. 2011). During chronic HIV infection, several MPS publications have demonstrated the therapeutic success of antivirals in the presence of a minority variant known to have resistance; whilst others have proposed that minority variants may have to be present at a threshold to become predominant (Delobel et al. 2011; Nicot et al. 2015).

4.1.3.3 Surveillance of vaccine formulations for genetic instability or a loss of immunogenicity

MPS technology is a putative replacement for the detection of minority adaptive mutations or reversions to virulence in live virus vaccine formulations (Neverov and Chumakov 2010). Several live RNA virus vaccines exist, which aim to eradicate disease or control seasonal infection and transmission. Two examples would be the

oral poliovirus vaccine (OPV) and the seasonal Influenza A vaccine. Prior to formulating a vaccine, Influenza A virus is replicated in embryonated chicken eggs and Poliovirus is replicated in mammalian cell cultures. MPS can detect mutations at high sensitivity which have occurred during Influenza A virus replication; mutations in the haemagglutinin protein can be beneficial to the vaccine development process and lead to an increase in replication efficiency during *in vivo* growth, or detrimental, leading to a decrease in vaccine antigenicity or safety in humans (Majid et al. 2015). Oral poliovirus (OPV) strains can revert to virulence, known as vaccine-derived poliovirus (VDPV), during *in vitro* replication (Sarcey et al. 2017). Reversion of a poliovirus vaccine strain is traditionally detected by Mutant Analysis by PCR and Restriction Enzyme Cleavage (MAPREC), however MPS can be equally effective and is not limited by the analysis of known sites of virulence only (Sarcey et al. 2017).

4.1.3.1 Surveillance of viruses in vaccinated populations

MPS can provide a powerful tool to monitor virus change in a population under vaccine pressure; high throughput and sequence depth are combined to monitor how a virus is changing in the population or elucidate conserved and variable epitopes to better inform on vaccine impact and potentially, vaccine efficacy and design.

Chong and colleagues recently applied MPS to Influenza A virus (IAV) isolates from individuals who had or had not received the vaccine, and observed an increase in the number of epitope AA changes, or positive selection, among viruses recovered from vaccinated individuals (Chong and Ikematsu 2017). Furthermore, MPS has

been used to identify VDPV minority variants in stool samples after administration of the OPV; unlike other studies, the surveillance of VDPVs was achieved without the potential bias of a tissue culture system (Chi et al. 2009; Sahoo et al. 2017; van der Sanden et al. 2009). The investigators noted MPS could detect VDPV reversion to neurovirulence at high sensitivity (< 1% frequency), which can occur if it is transmitted amongst a community or through chronic infection in immunocompromised individuals (Sahoo et al. 2017).

MPS is also an integral part of a recently published technology which demonstrated an ability to isolate single HIV particles with a broadly neutralizing antibody, and subsequently generate envelope sequence data to aid in the identification of vaccine candidates (Chaipan et al. 2017).

4.1.4 True variant calling vs error rate

Four out of the six MPS platforms currently available (Bridge amplification, Pyrosequencing, Semiconductor sequencing and Single molecule real-time sequencing) determine nucleotide base calls by DNA polymerase incorporation and strand elongation, and cannot sequence RNA directly (Bentley et al. 2008; Braslavsky et al. 2003; Margulies et al. 2005; Rothberg et al. 2011). Therefore, an RNA virus genome must be reverse transcribed to cDNA, and can require PCR amplification prior to MPS if the amount of RNA before, or cDNA after, RT is low. From sample preparation to data analysis, cumulative errors can be induced by enzyme misincorporations at each step.

The threshold for distinguishing true minority variants from erroneous base calls on the Genome Analyzer IIx platform (Illumina) was found to be ≥ 2000-fold coverage

at a base position, for which the true mutant base call represented ≥ 1% (Thys et al. 2015). This threshold was determined by sequencing a HCV plasmid mixture containing a wild type and a mutant at different ratios (Thys et al. 2015). However inter-run variability can occur (Verbist et al. 2015). Most MPS investigations which aim to detect the presence of minority variants will amplify and sequence a plasmid, containing the virus genome, to indicate how much PCR and the MPS platform contribute to background error (Bartolini et al. 2013; Delobel et al. 2011; Van den Hoecke et al. 2015; Nasu et al. 2011; Nicot et al. 2015; Sarcey et al. 2017). Furthermore, if errors generated during MPS are random, the reliability of observed minority variants can be increased by performing sample replicates (Majid et al. 2015). Otherwise, true and false positives can be discerned by filtering variants based on thresholds set in the literature or several criteria which, in theory, should separate error from true base calls (Batty et al. 2013; Hasing et al. 2016).

4.1.5 Selective pressure on HuNoV in an immunocompetent or immunocompromised host

Accurately identified HuNoV minority variants in a host can be measured to indicate whether the population is undergoing selection. HuNoV, in a similar manner to other RNA viruses, will undergo positive selection to increase fitness of the virus population, or purifying selection to maintain fitness (Domingo et al. 1997). The selective pressures placed upon a HuNoV population are an avoidance of neutralization by the host immune response, maintenance of virus ability for transmission between hosts and through environmental routes, and to preserve or enhance host receptor binding affinity.

4.1.5.1 Chronic infection

In the literature, MPS studies of HuNoV chronic infection have largely investigated individuals awaiting or following bone marrow transplantation (BMT) (Hasing et al. 2016; Kundu et al. 2013; Vega et al. 2014b). Prior to BMT, an individual must be highly immunosuppressed to permit acceptance of the graft and prevent graft-versus-host disease, and thereafter immune reconstitution is stringently controlled (Welniak et al. 2007).

Bull and colleagues were the first to use MPS, the 454 (Roche) instrument, to characterise chronic HuNoV infection by targeting the ORF2/3 region in an infant with a severe undefined immunodeficiency (Bull et al. 2012). Over the course of infection (Day 1, 4 and 288) a highly heterogeneous population of HuNoV minority variants were present at low frequencies. Subsequently, minority variants in 18 bases (3/5 non-synonymous mutations in epitopes of the P2 domain) reached fixation in the last time point that were not observed in the early stages (Day 1 or 4) of infection (Bull et al. 2012). A separate study using similar MPS technology in three chronic HuNoV infections (random sampling across days 5-19, 1-61 and 0-30, respectively), of patients which were awaiting or had undergone BMT, identified an excess of synonymous mutation fixation (846/1082), implying the presence of purifying selection (Kundu et al. 2013). In the ORF2 region, 50% of the non-synonymous mutations were present in the P2 domain. However, within the P2 subdomain the number of synonymous mutations exceeded the non-synonymous mutations (67 versus 40), and none of the AA changes corresponded to known epitopes (Kundu et al. 2013). In contrast, Vega and colleagues sampled chronic

infection in three BMT patients at single time points and described the presence of positive selection in the P2 region by Sanger sequencing, when VP1 sequences were compared to GII.4 HuNoVs circulating in the population (Vega et al. 2014b). Furthermore, MPS of the whole virus genome in the same samples, via the 454 (Roche) system, demonstrated most variable sites and positively selected sites were in the VP1 gene (Vega et al. 2014b).

4.1.5.2 Acute infection

MPS studies solely characterising minority variants in acute HuNoV infection are underrepresented in the literature. In an immunocompetent host, HuNoV minority variants have been reported as much more homogenous. Over the course of infection (Day 1 and 10), a single minority variant was detected (~20 %) in the ORF2/3 region and the maximum difference in variant frequency between both time points was 2.6% (Bull et al. 2012). This finding agrees with the MiSeq (Illumina) findings of Hasing and colleagues, who were unable to identify any minority variants in four patients with acute HuNoV infection at single time points (Hasing et al. 2016).

4.1.6 The role of copro-antibody in HuNoV immunity

The correlates of HuNoV protection reported so far are host genetics, serum HBGA blocking antibody (Ab), serum haemagglutination inhibition Ab, salivary immunoglobulin A (IgA), faecal IgA, virus-specific memory immunoglobulin G (IgG) cells, serum HBGA blocking Ab and serum IgA (Ramani et al. 2016). Several of these are components of the adaptive immune system, and could place selective pressure upon HuNoV during infection.

The site of HuNoV infection in the human host is the small intestine (de Graaf et al. 2016), and therefore mucosal-antibody should have an important role in virus neutralisation. The appearance of HuNoV strain-specific faecal IgA in an infant has been demonstrated to coincide with a reduction in genome copy numbers, and eventual clearance (Iritani et al. 2007). However, in adult volunteer studies, NV-specific faecal IgA has not been found to neutralise infection or illness, but did correlate with reduced viral loads and a decreased duration of virus shedding (Okhuysen et al. 1995; Ramani et al. 2015). In a similar manner to that described by Iritani, copro-antibodies are described as having an important role in the protection of infants from Rotavirus (RoV) infection, and were recently shown to correlate with serum IgA responses pre- and post-vaccination (Lappalainen et al. 2017). Coulson and colleagues described the existence of a RoV specific copro-antibody plateau in a subset of infants, after repeated natural virus infection, which reduced the number of symptoms or symptom severity upon infection (Coulson et al. 1992). The presence of a RoV specific copro-antibody plateau paralleled virus-neutralising Ab in faeces to several RoV serotypes (RV 3, RV 4, Wa, Ku and ST-3) (Coulson et al. 1992).

4.2 Materials and methods

4.2.1 Clinical sample preparation, enrichment, library preparation and sequencing

Faecal suspensions were prepared from archived, and newly collected, longitudinal samples from six patients (A-F) (Table 4.1). Ethical approval for collection and the method of preparation was identical to the protocol described in Chapter 3. The

stool suspension supernatant was enriched with PGM-MBs (10:1 by volume), at pH 3.6 and room temperature (RT) for 15 minutes as previously described. The PGM-MBs were used to capture and pool each sample 3 to 10-fold, depending on the amount of stool available.

Material captured by the bead-protein complex was magnetically separated from the supernatant, and the nucleic acid extracted with a Gn silica nucleic extraction (Boom et al. 1990). Subsequent post-extraction methodology and library preparations were identical to the protocols described in Chapter 3.

MPS was performed in an identical manner to Chapter 3, however, library preparations from patients A, B and C sequences were prepared in duplicate rather than once (Table 4.1). In total two sets of library preparations were prepared for each sample. The first set of library preparations were sequenced on flow cell 1 (Table 4.2). The sequencing process was then repeated for set 1 in combination with set 2 on flow cell 2 (Table 4.2).

Table 4.1 Quantification of GII HuNoV in archived and collected clinical samples by qPCR (One case was severely immunocompromised (B) whilst no known immunodeficiencies or immunosuppressive therapy were present in the remaining five cases).

Patient ID	Immunosuppresive comorbidity	Genotype	Year of detection	Day collected post recruitment	GII HuNoV genome copies/qPCR reaction
A	None	GII.4	2012	3	7.59×10^{7}
				7	2.17×10^{7}
				15	1.33×10^{3}
B	Haemopoetic stem cell transplantation	GII.7	2012	3	1.03×10^{9}
				9	1.29×10^{9}
				12	1.54×10^{8}
				15	1.29×10^{9}
C	None	GII.4	2012	3	3.23×10^{7}
				5	2.94×10^{7}
				6	1.22×10^{7}
				10	5.91×10^{5}
				11	1.25×10^{4}
D	None	GII.4	2017	1	4.36×10^{5}
				2	3.59×10^{5}
E	None	GII.16P/GII.4	2017	0	7.11×10^{6}
				1	1.41×10^{7}
				1	9.15×10^{6}
F	None	GII.4	2017	0	1.69×10^{5}
				1	2.40×10^{7}
				2	6.89×10^{6}
				3	1.46×10^{6}

Table 4.2 The order of MPS replicates for patients A, B and C.

Patient ID	HiSeq (Illumina) flow cell number	
	1	2
A, B and C	Replicate 1	Replicate 1
		Replicate 2

4.2.2 Expression of HuNoV GII Virus-like particles (VLPs)

Two HuNoV positive stool samples, containing GII.4 New Orleans and GII.4 Sydney strains, were prepared as a 10-20% (v/v) faecal suspension in Medium 199 (Fisher

Scientific, Leicestershire, UK), and the nucleic acid extracted by Gn silica method (Boom et al. 1990).

Copy DNA of the ORF1+ORF2 region was generated with the SuperScript™ III Reverse Transcriptase System (Fisher Scientific, Leicestershire, UK). Firstly, 40 μL of the nucleic acid extract and 0.9 μM of primer TVN (Table 4.3) were denatured at 70°C for 5 minutes, and incubated on ice for 2 minutes. The RT reaction mixture (28 μL) was combined with the extract after denaturation, and thus the concentrations overall (68 μL in total) were 1 × First-Strand Buffer, 5 mM $MgCl_2$, 0.5 mM dNTPs (Fisher Scientific, Leicestershire, UK), 200 units of SuperScript™ III reverse transcriptase/reaction. The RT reaction was incubated at 42°C for 1 hour and then on ice for 2 minutes.

The cDNA was amplified with the Expand High Fidelity System (Roche, West Sussex, UK) to amplify from the ORF1/2 junction to the 3' untranslated region (UTR). The PCR reaction comprised 1 × Expand High Fidelity Buffer with $MgCl_2$ (Roche, West Sussex, UK), 0.5 mM dNTPs (Fisher Scientific, Leicestershire, UK), 2.5 mM $MgCl_2$ (Roche, West Sussex, UK), 0.4 μM primer ORF1/2-F1, 0.4 μM primer TVN-Linker and 3.5 units of Expand High Fidelity Enzyme Mix (Roche, West Sussex, UK) in a total volume of 50 μL. The amplification mixture underwent a Touch-Down PCR thermal profile of 95°C for 5 minutes followed by three cycles of denaturation at 95°C for 1 minute, annealing at 60°C for 30 seconds and extension at 72°C for 3 minutes. After cycling the annealing temperature decreased incrementally by 2° until 50°C, at which point 25 cycles would occur, prior to a final extension step of 72° for 5 minutes.

Each amplicon of the expected size was purified from agarose gels with the Minelute Gel Extraction kit (Qiagen, West Sussex, UK), following the manufacturer's instructions. The purified amplicons were cloned into the intermediate vector pCR2.1-TOPO® (Fisher Scientific, Leicestershire, UK), and modified for directional cloning (Appendices D and E) into a vector containing overlapping regions with *Autographa californica nuclear polyhedrosis virus* (AcMNPV) for lipid mediated co-transfection and purification (Appendices F, G and H), as described previously (Allen et al. 2009). The protein concentration of the VLP preparation was measured using the Qubit Protein Assay Kit (Fisher Scientific, Leicestershire, UK).

Table 4.3 Primer oligonucleotides.

Primer	Sequence (5'-3')	GC[a] (%)	Tm[b] (°C)	Amplicon size (bp)	Reference
ORF1/2-F1	CTG AG CAC GTG GGA GGG CG	74	59.7	~2500	(Allen et al. 2008)
TVN-Linker	CGA CCT AGG TGA TAC ATG AT	45	49.7		(Allen et al. 2009)
TVN	NTT TTT TTT TTT TTT TTT TTT CGA CCT AGG TGA TAC ATG AT	22-24	57.5-58.5		
New Orleans modification 1	AGA TAT CGA GCT CTA TAA ATA TGA AGA TGG CGT CGA GTG ACG	43	66.5	2500	This study
New Orleans modification 2	AGA TAT CGC ATG CTT TTT AAA AGA CAT CAG AGA AAA AGA AAG ATA A	28	61.9		
Sydney modification 1	AGA TAT CGG ATC CTA TAA ATA TGA AGA TGG CGT CGA GTG ACG C	44	67.4	2501	
Sydney modification 2	AGA TAT CCT GCA GTT TTT AAA AGA CAC TAA AGA AAA AGA AAG ATA A	26	61		

[a] **Guanidine / Cytosine content.**

[b] **Melting temperature (20 µM - Basic).**

4.2.3 Serology

4.2.3.1 Sandwich enzyme-linked immunosorbent assay (ELISA) for the detection of total faecal IgA

Flat-bottom Costar Stripwell™ 96 well plates (Sigma, Dorset, UK) were coated with 100 µL of polyclonal rabbit anti-human IgA (Stratech Limited, Suffolk, UK), diluted, 1 in 4000 in carbonate-bicarbonate buffer, pH 9.6 (Sigma, Dorset, UK). Each plate containing coating buffer was incubated overnight in a moist environment at 4°C.

All subsequent incubation steps were performed at 37°C in a moist environment. After overnight incubation, coated plates were washed five times with 300 µL of wash buffer, which consisted of sterile PBS and 0.05 % (v/v) TWEEN 20 (Sigma, Dorset, UK). Each well was then blocked in 300 µL sterile PBS (Sigma, Dorset, UK) and 5% (w/v) instant dried skimmed milk, pH 7.4, for 2 hours. Faecal suspensions were clarified by centrifugation, as described previously, and 100 µL of each serial dilution (1:50, 1:500 and 1:1000) incubated for 2 hours on the plate. Prior to detection each well was washed with 300 µL of wash buffer three times. The total IgA bound was detected by adding 100 µL polyclonal goat-anti human IgA conjugated to horse radish peroxidase (BioRad, Herfordshire, UK), diluted 1 in 4000 in sterile PBS (Sigma, Dorset, UK) and 1% instant dried skimmed milk, for 1 hour. The plate was then washed five times with 300 µL of wash buffer, and subsequently, a signal was developed by adding 100 µL 3',3',5',5'-tetramethylbenzidine liquid substrate (TMB), supersensitive, for ELISA (Sigma, Dorset, UK) to each well prior to a 10-minute incubation in darkness. Once the incubation period was finished, the reaction was stopped with 50 µL of 1 M H_2SO_4

and the absorbance read at 450 nm (A_{450nm}) on the Multiskan® Spectrum (Fisher Scientific, Leicestershire, UK) microplate reader. A dilution series of a known concentration of Purified Human Secretory IgA (BioRad, Herfordshire, UK) and three negative controls (faecal suspension replaced by sterile PBS) were included in each assay.

4.2.3.2 Capture ELISA for the detection of strain specific IgA

Strain specific IgA was determined in an identical manner to that described above with the following exceptions. Firstly, plates were coated with 7.5 mg/mL PGM III (Sigma, Dorset, UK). Secondly, wells were blocked for 1, rather than 2 hours, and after the blocking stage washing was repeated three times in total. Thirdly, prior to the addition of serially diluted sample, a preparation of partially purified VLP was added for 1.5 hours, and washed three times with wash buffer (Farmington Hills = 20 μg/well, New Orleans or Sydney = 2 μg/well). Lastly, after the addition of faecal suspension, the plate was incubated for 1.5, rather than 2 hours.

4.2.4 Data analysis

4.2.4.1 Bioinformatic analysis

HiSeq (Illumina) sequences were analysed in an identical manner to that described in Chapter 3. To identify a true consensus sequence change or minority variant at one time point, in patient A, B or C, complete agreement between all library preparation replicates was necessary. The frequency of true minority variants at different coverage levels was then deduced, and used as a threshold to analyse each time point in patients D, E and F.

Mutations which altered the molecular surface of the HuNoV P domain were mapped with YASARA (www.yasara.org - version 16.11.20) and modelled with NOC (noch.sourceforge.net - version 3.01). The ORF2 of selected consensus sequences were modelled against the HuNoV strain NSW0514, and is available as a PDB file from www.rcsb.org, DOI:10.2210/pdb4oos/pdb (Singh et al. 2015).

4.2.4.2 Phylogenetic analysis

Phylogenetic analysis was performed with Molecular Evolutionary Genetics Analysis (MEGA) software (version 7). Nucleic acid sequences were aligned with Clustalw, under the default settings, and phylogenetic relationships were inferred with the Maximum-Likelihood algorithm (Kumar et al. 2016; Tamura and Nei 1993). Phylogenetic trees were subsequently reformatted using the online Interactive Tree of Life (ITOL) software (Letunic and Bork 2016).

4.2.4.3 ELISA data analysis

A standard curve of a known concentration of pure secretory human IgA was generated by four parameter logistic regression ("Four Parameter Logistic Curve" online data analysis tool, MyAssays Ltd., 6th September 2017, http://www.myassays.com/four_parameter-logistic-curve.assay). The total IgA and VLP specific IgA concentration was determined by subtracting sample A_{450} from the average of the negative control replicates. A sample of $A_{450} \leq 0.3$ was negative, and those above were extrapolated against the standard curve to deduce VLP specific IgA as a percentage of total IgA.

4.2.4.4 Statistical analysis

Minority variant distributions were analysed by a Wilcoxon rank sum test.

4.3 Results

4.3.1 PGM-MB enrichment of longitudinal samples was reproducible and the mean coverage reflected the stage of clinical infection

HiSeq (Illumina) reads from patients A, B, C, D, E and F were assembled into contiguous sequences *de novo*, and a local nucleotide alignment showed all but one infections were GII.4 HuNoVs, whilst the virus present in patient B was a GII.7 HuNoV (Table 4.1). A consensus was derived from the highest quality replicate from each patient for read alignment, or the earliest time point if replicates were not available, and this meant read coverage and the percentage of the HuNoV genome recovered could be determined (Tables 4.4 - 4.7).

Table 4.4 Summary of GII.4 virus reads recovered from the longitudinal samples of patient A (* = Failed library, ** = Unusually high number of total reads and IQR=Interquartile range).

Day	Replicate	Coverage		Percentage of genome recovered (%)
		Median	IQR	
3	1	129	269	99.6
	2**	3634	2553	100.0
	3	2	2	27.9
7	1*	0	NA	0.0
	2	5	6	83.7
15	1	2	2	22.3
	2	2	1	26.4
	3	1	1	9.5

Table 4.5 Summary of GII.7 virus reads recovered from the longitudinal samples of patient B (*= Failed library and IQR=Interquartile range).

Day	Replicate	Coverage		Percentage of genome recovered (%)
		Median	IQR	
3	1	14	13	97.7
	2	54	43	99.8
	3	86	86	99.8
9	1	45	35	99.9
	2	56	47	99.9
	3	53	49	100.0
12	1	189	136	99.8
	2	1	0	0.3
	3	13	12	98.9
15	1	40	28	99.8
	2	161	131	99.8
	3	41	38	99.8

Table 4.6 Summary of GII.4 virus reads recovered from the longitudinal samples of patient C (IQR=Interquartile range).

Day	Replicate	Coverage		Percentage of genome recovered (%)
		Median	IQR	
3	1	136	88	99.8
	2	5	4	80.8
5	1	323	201	100
	2	513	294	99.9
	3	8	9	94.7
6	1	139	91	99.9
	2	28	17	99.2
	3	892	551	99.9
10	1	5	7	77.8
	2	10	11	96.8
11	1	1	1	13.0
	2	3	4	73.8
	3	2	1	49.1

Table 4.7 Summary of GII.4 virus reads recovered from the longitudinal samples of patient's D, E and F (IQR=Interquartile range).

Patient ID	Day	Coverage		Percentage of genome recovered (%)
		Median	IQR	
D	0	36	38	99.64
	1	53	51	99.3
E	0	78	63	99.9
	1	117	122	100.0
	1	43	310	99.8
F	0	201	166	100.0
	1	16	17	99.1
	2	7	7	86.8
	3	5	7	88.5

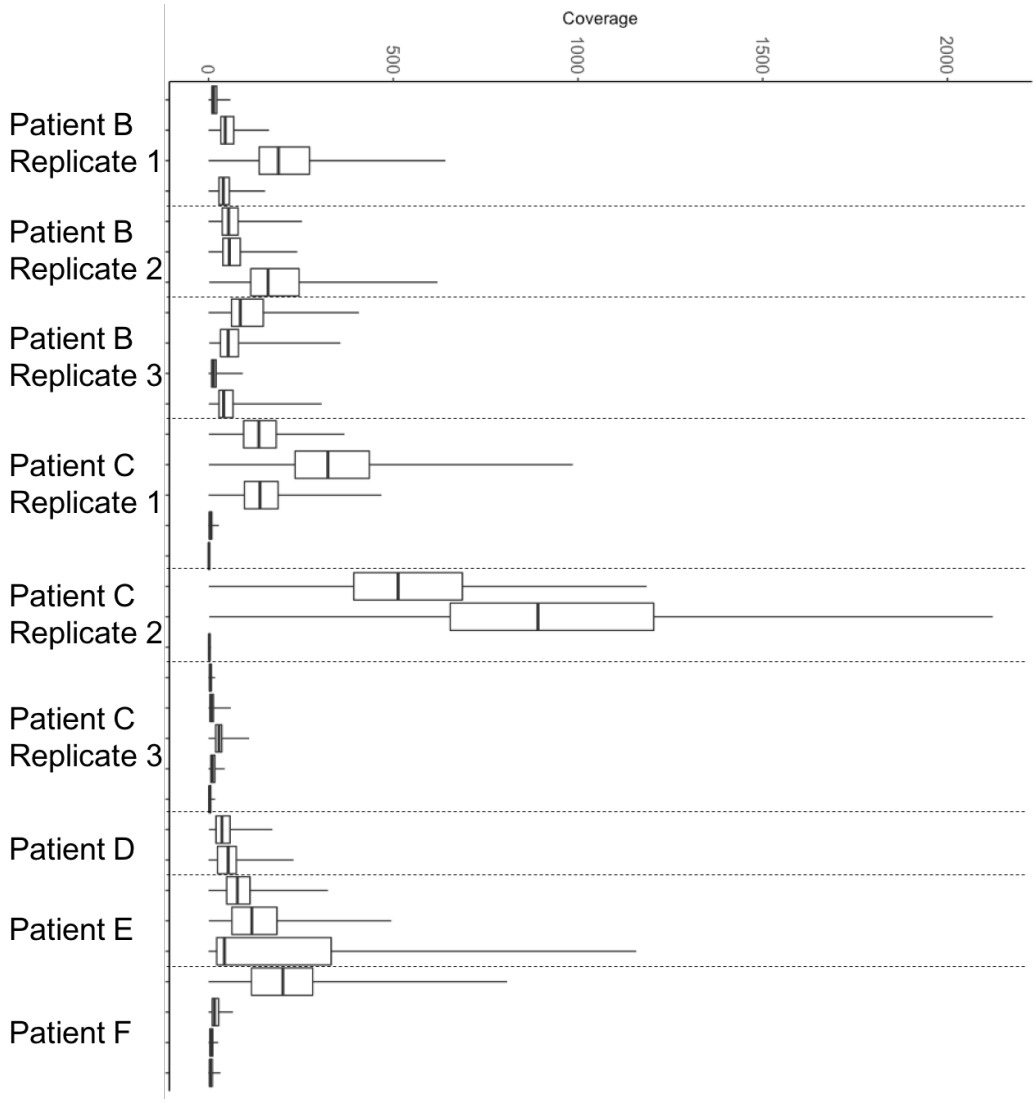


Figure 4.1 The range, interquartile range (IQR) and median of coverage for each sample analysed.

If library preparation failures were excluded, complete or near complete genomes were recovered, except for the third replicate collected on day 3 from patient A and the latter time points of patient A (Day 15) and C (Day 11). Patient A and C were the only two immunocompetent individuals who had samples collected for longer than a week, and at the latter time points partial HuNoV genomes were recovered at low median coverage (Tables 4.4 and 4.6). The decrease in median coverage and loss in genome recovery coincided with a low virus titre on day 15 and 11 for patient A and C, respectively (Tables 4.2, 4.4 and 4.6). This correlation however could not be seen

in patient B or D, E and F, and was most likely due to chronic illness and a shorter clinical sample collection period (Tables 4.5 and 4.7). In contrast, patient B had consistently high virus titres (Table 4.2) and, despite variability in coverage between replicates, the median coverage was at least tenfold across the HuNoV genome (Table 4.5). In the thirty-three clinical samples analysed (Patient's B - F), twenty-nine had an upper quartile range greater than the lower quartile range (Figure 4.1).

4.3.2 The GII HuNoV consensus genome was highly stable in acute infection

In total, three replicate data sets were available for most of the samples from patient A and C, and all the samples collected from patient B (Tables 4.4 - 4.6), therefore consensus genome changes could be inferred reliably. However, due to an insufficient amount of longitudinal sequence data from patient A none of the replicates were analysed, whilst one replicate from patient B (Day 12) and three replicates from patient C (one each from day 3, 10 and 11) were excluded (Appendix C). In patient's B and C, three synonymous consensus genome changes were detected in total over the course of both infections (Tables 4.8 and 4.9).

Table 4.8 Consensus genome changes observed in longitudinal samples from Patient B.

Patient B					
Day of sample collection post recruitment	Replicate	Nucleotide frequency at position (%)			
		725			
		A	C	G	T
3	1		25		75
	2		45		55
	3		43.1		56.9
9	1		60.6		39.4
	2		43.5		56.5
	3		61.3		38.7
12	1		64.1		35.9
	2		33.3		66.7
15	1		66.7		33.3
	2		55.8		44.2
	3		72.7		27.3

Table 4.9 Consensus genome changes observed in longitudinal samples from Patient C.

Patient C									
Day of sample collection post recruitment	Replicate	Nucleotide frequency at position (%)							
		5323				5431			
		A	C	G	T	A	C	G	T
3	1				100			100	
	2				100			100	
5	1				100			100	
	2	0.2	0.2		99.6			99.8	
	3				100			100	
6	1				100			100	
	2		0.2		99.8	0.1	0.1	99.8	
	3		9.8		90.2	3.6		96.4	
10	1		70		30	21.1		78.9	
	2		0		100	0		100	
11	1		100			100			
	2		100			100			

The synonymous mutation in patient B was in position 725 of the p48 gene of ORF1 (Table 4.8). The sequence data obtained over all four time points indicated that the alleles at position 725 were present as a heterogeneous mixture of cytosine and

thymine. For patient B, on day 3 most of the alleles at position 725 were thymine, and then on day 9 and 12 a mixture of cytosine and thymine existed. However, by day 15 most of the alleles in each replicate were cytosine (Table 4.8).

The two synonymous mutations detected in patient C were present at positions 5323 and 5431 of the ORF2 S domain (Table 4.9). In contrast to patient B, these consensus alleles were highly homogenous. For patient C, on day 3, 5 and 6 the consensus allele was thymine and guanine at position 5323 and 5431, respectively (Table 4.9). Although minority alleles were detected on day 6, the allele switching point for position 5323 (T > C) and 5341 (G > A) was between day 10 and 11 and subsequently became fixed on day 11 (Table 4.9).

4.3.3 High quality base call filtering led to a marginal loss of sequence data

In HiSeq (Illumina) read data, every base call is assigned a Phred quality score, which the instrument determines as the probability a nucleotide is incorrect. The Phred score can be calculated as the -10 $\log_{10}$ of the error probability, and therefore a score of 30 would mean the inferred base is 99.9% likely to be correct (Ewing and Green 1998). To ensure minority variants were being called with high quality data sets, base calls with a Phred score < 30 were filtered out.

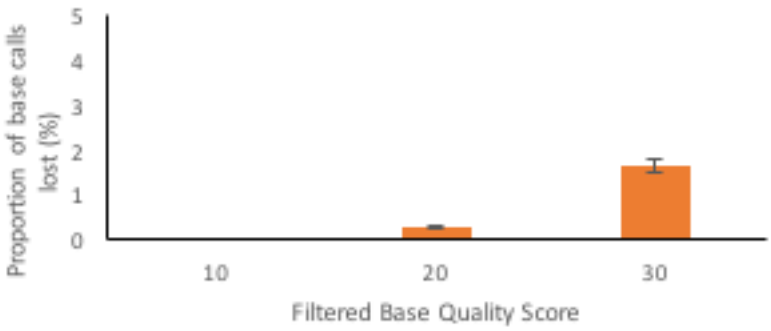


Figure 4.2 The effects of Phred score filtering on HuNoV clinical sample data sets.

The data sets were filtered at high and low Phred score stringency to examine how much data would be lost. If base calls below Q30 or Q20 were removed 1.6 % or 0.23 % of the total nucleotides in the sequence data were lost (Figure 4.2).

4.3.4 The distribution of minority variants across different codon positions were more evenly distributed in acute GII.4 infection

To investigate whether any selective pressure existed upon the virus populations present in patient B and C, and in which regions of the HuNoV genome this occurred, the presence of minority variants was measured at the 1st, 2nd and 3rd codon positions of each AA reading frame. Most AA redundancy exists in the 3rd codon position, and therefore could be used as a proxy to determine the presence of positive or purifying selection. For this analysis, one replicate from patient B was excluded due to library failure (Day 12 replicate 2), two replicates from the early time points of patient C were excluded due to library failure or low coverage (Day 3 replicate 2 & day 5 replicate 3) and all the latter time points from patient C due to low coverage (Day 10 and 11) (Tables 4.5 and 4.6).

In chronic GII.7 infection (patient B), a disproportionately larger number of minority variants were present in the 1st codon position of ORF1 over all time points (Figure 4.3). In relation to ORF1, ORF2 and ORF3 contained few minority variants which were at similar frequencies across the three codon positions, however, there were two exceptions (Figure 4.3). The exceptions existed in ORF3, where a larger frequency of 2nd and 3rd codon mutations were detected on day 9 (replicate 1) and a similar pattern was present on day 11 (replicate 1) at the 1st codon position (Figure 4.3). On the other hand, in acute GII.4 infection (Patient C) the frequency of variants appeared to decrease proportionally across all ORFs, the 1st codon position had the most minority variants, followed by the 2nd and 3rd codon positions, respectively (Figure 4.3). Interestingly, an increase in the frequency of minority variants over time, whilst coverage remained consistent, was observed in patient C, between day 3 replicate 1 and day 6 replicate 1 (Figure 4.3).

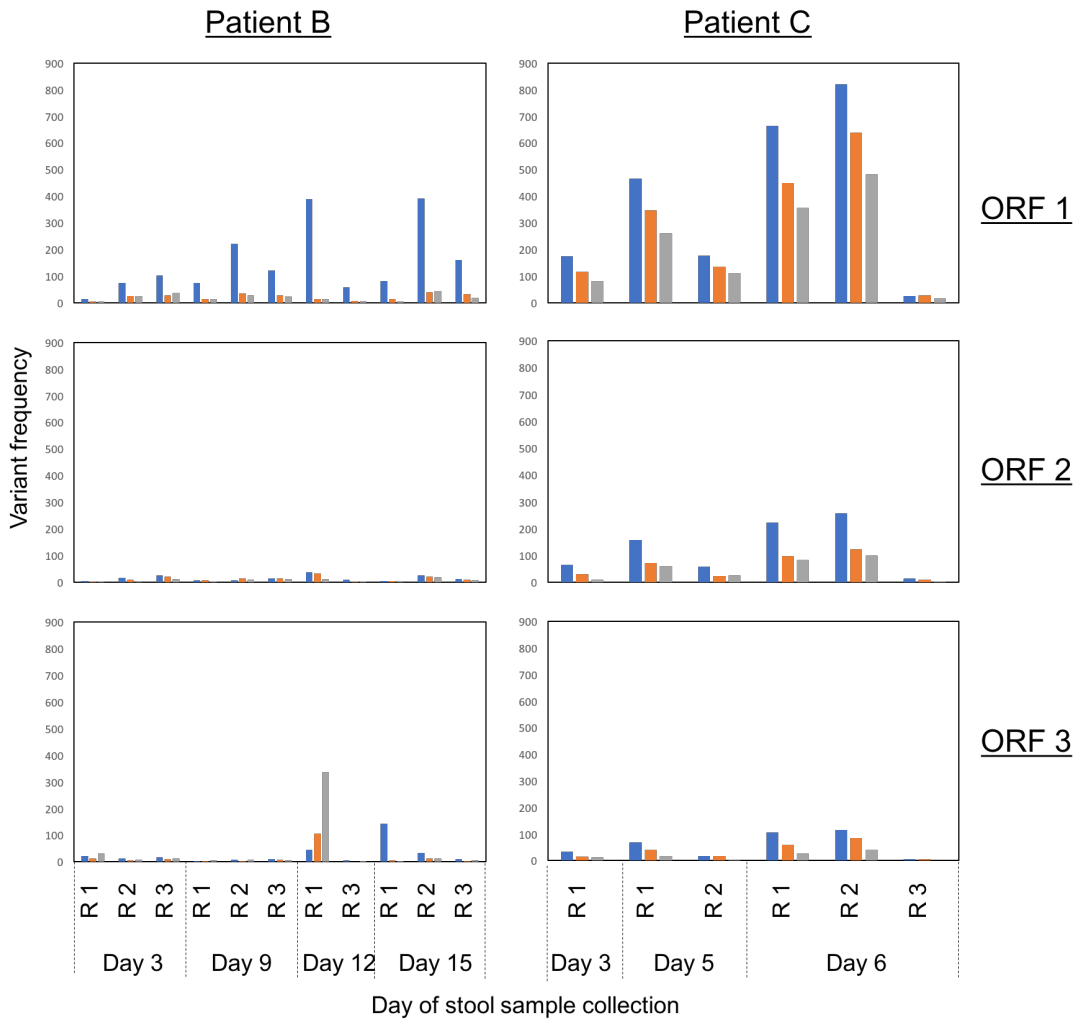


Figure 4.3 Minority variant frequencies in longitudinal samples of patient B and patient C at different codon positions (Blue = position 1, Orange = position 2, Grey = position 3, R = replicate).

If the minority variant frequencies at each codon position were normalised by gene size there were no differences between the ORFs in patient C, except less 3rd position mutations were present in ORF3 compared to ORF1 (Table 4.10). In patient B, however, the 1st codon position had more mutations in ORF1 ($p = < 0.05$) and 3 ($p= < 0.05$) when compared to ORF2. In the 2nd codon position there were more in ORF2 and 3 ($p= < 0.05$) than ORF1 and in the 3rd position more were present in ORF3 than 1 ($p= < 0.05$) or 2 ($p= < 0.05$) (Table 4.10).

Table 4.10 The mean minority variant frequencies normalised by ORF size in patient B and C.

Patient	Codon position	ORF		
		1	2	3
B	1	0.029±0.0076	0.0087±0.0019	0.040±0.017
	2	0.0040±0.00070	0.0073±0.0017	0.022±0.013
	3	0.0037±0.0011	0.0043±0.001	0.056±0.002
C	1	0.076±0.025	0.080±0.025	0.070±0.023
	2	0.056±0.019	0.037±0.011	0.045±0.015
	3	0.043±0.014	0.029±0.010	0.020±0.0075

4.3.5 GII.4 acute infection minority variants were present in more genes across the whole HuNoV genome

Investigations were performed to identify the specific locations of mutations, and whether minority variants persisted in chronic GII.7 (Patient B) and acute GII.4 (Patient C) infection. If a minority variant was present in all the replicates of a single time point, the genomic position was carried forward for further analysis, and if this variant crossed over into a later time point, it was considered persistent.

The minority variants present in the GII.7 virus population existed in less positions across the whole genome, but of those detected most were present at all four time points from patient B when compared to the GII.4 infection in patient C (Figures 4.4 - 4.8). However, this could be due to the analysis of more replicates and time points in patient B, which meant more stringent criteria for minority variant filtering.

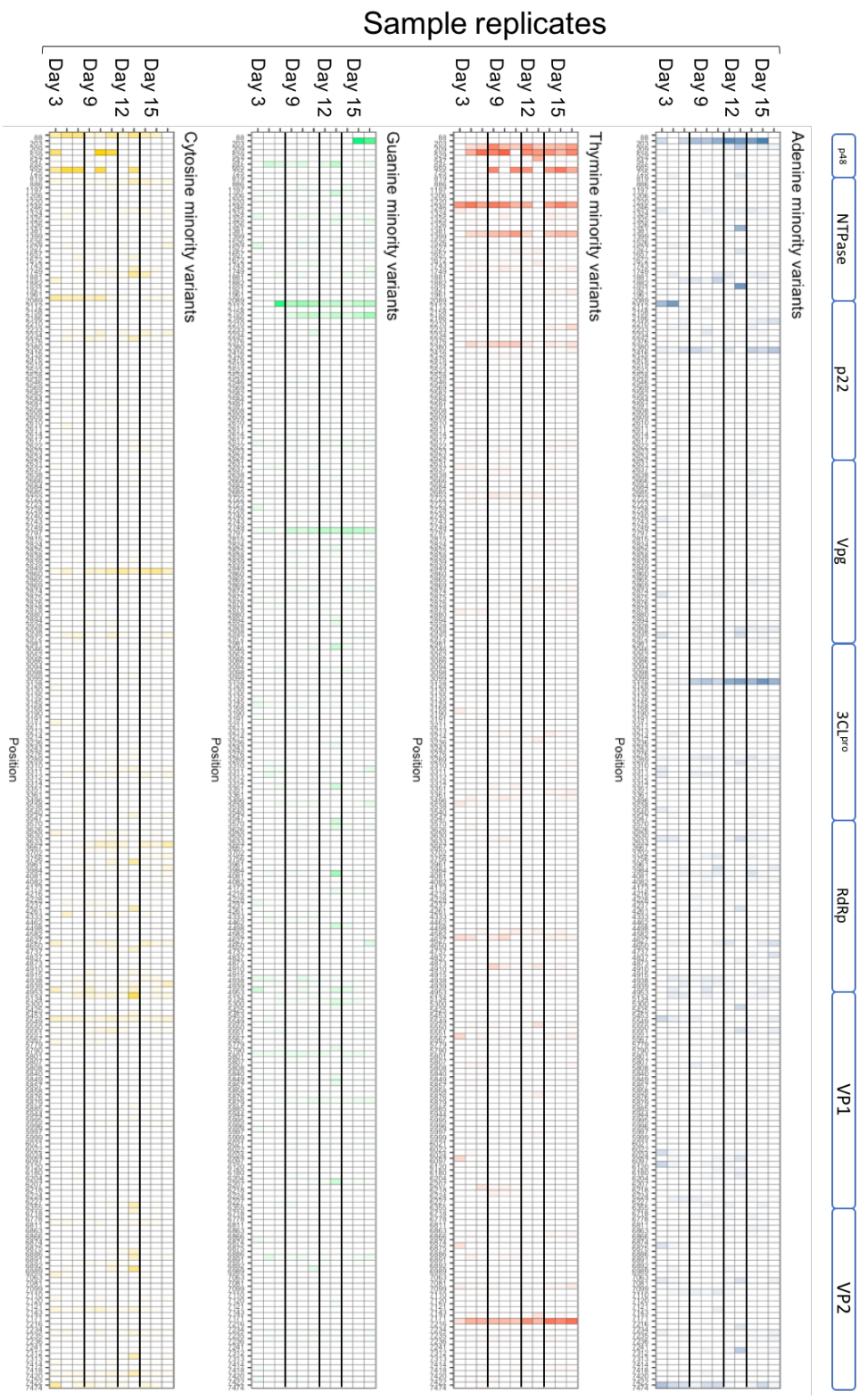


Figure 4.4 Genome locations of minority variants from day 3 to 15 of chronic GII.7 HuNoV infection in patient B.

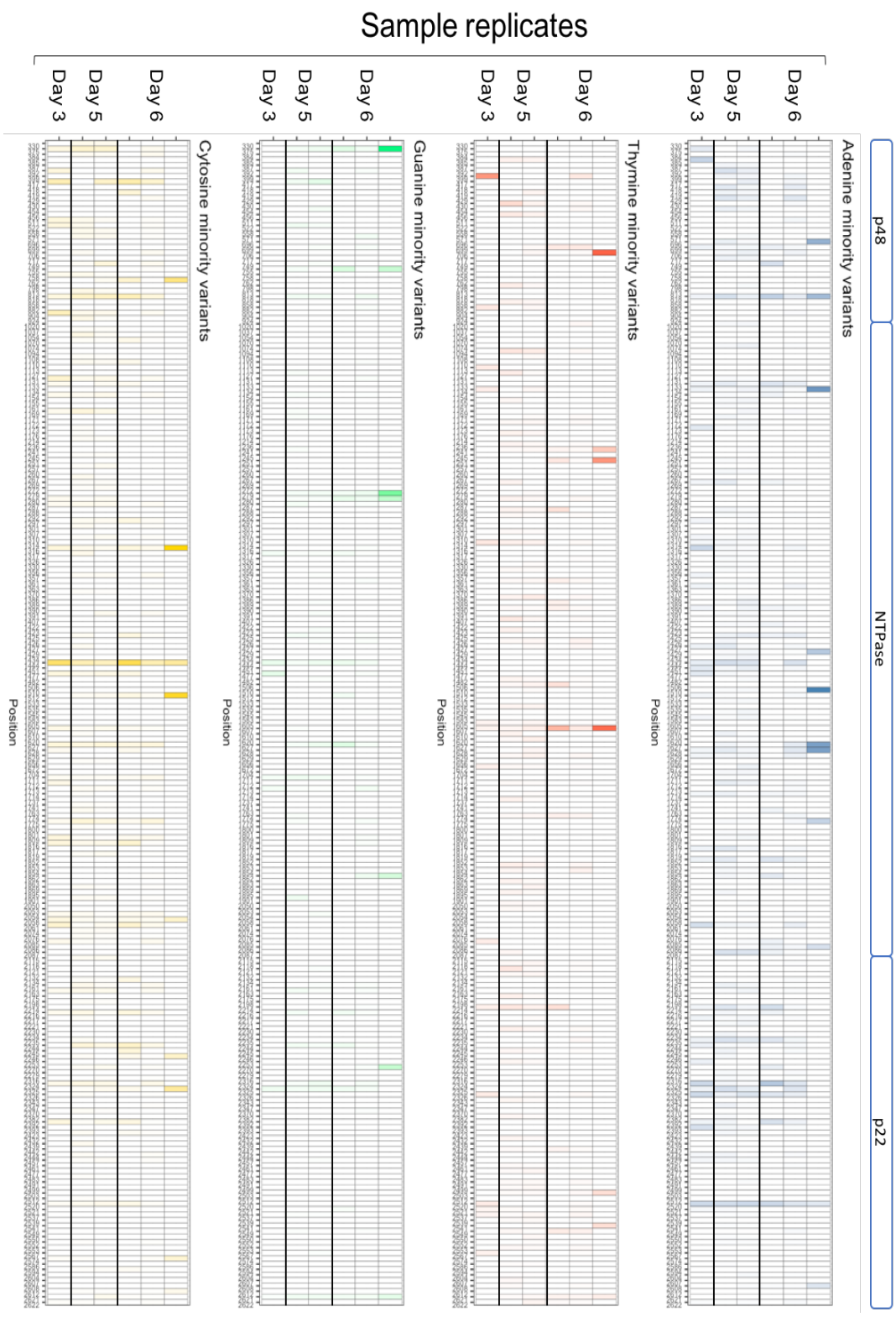


Figure 4.5 Minority variants located in the p48, NTPase and p22 non-structural genes over the course of acute GII.4 infection in patient C.

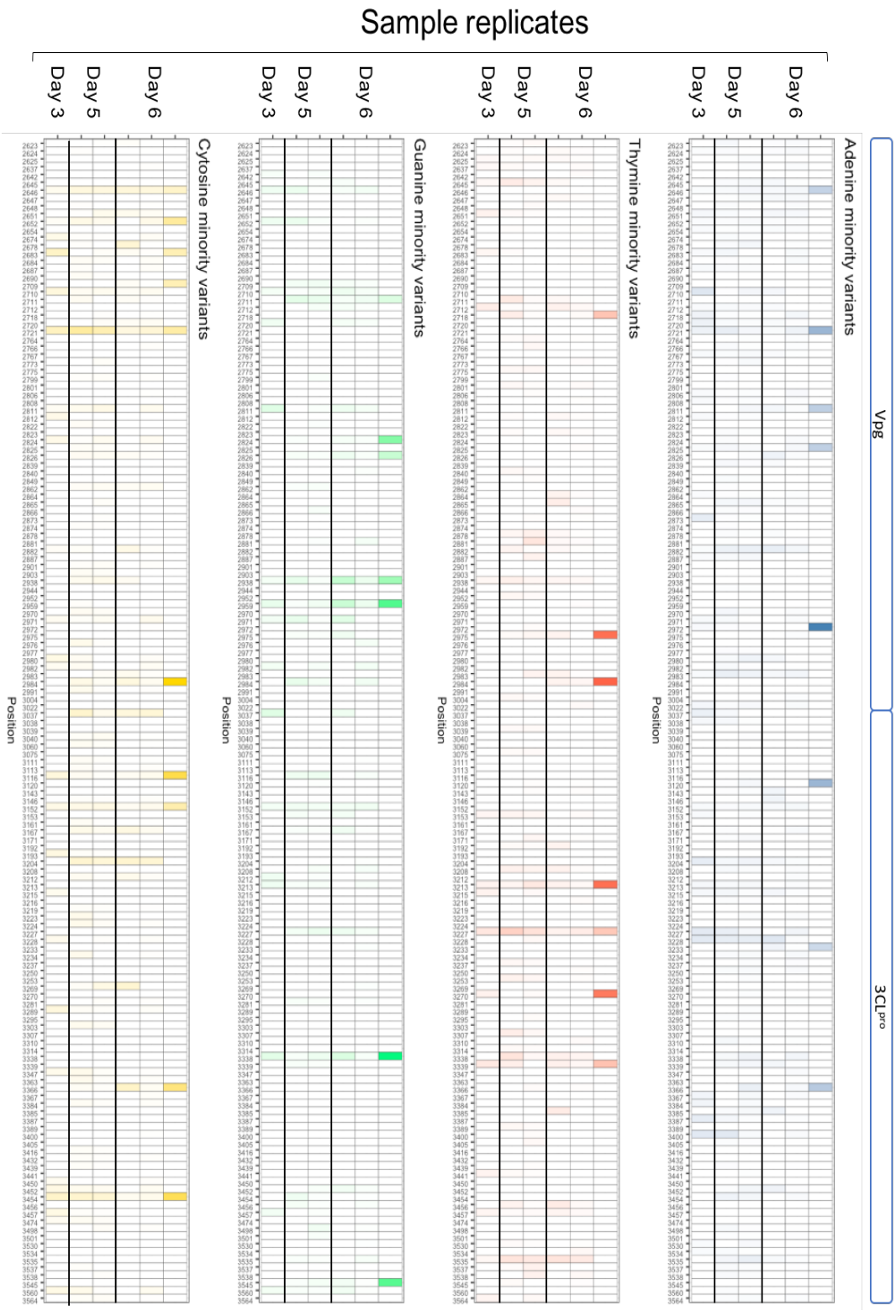


Figure 4.6 Minority variants located in the Vpg and 3CL protease non-structural genes over the course of acute GII.4 infection in patient C.

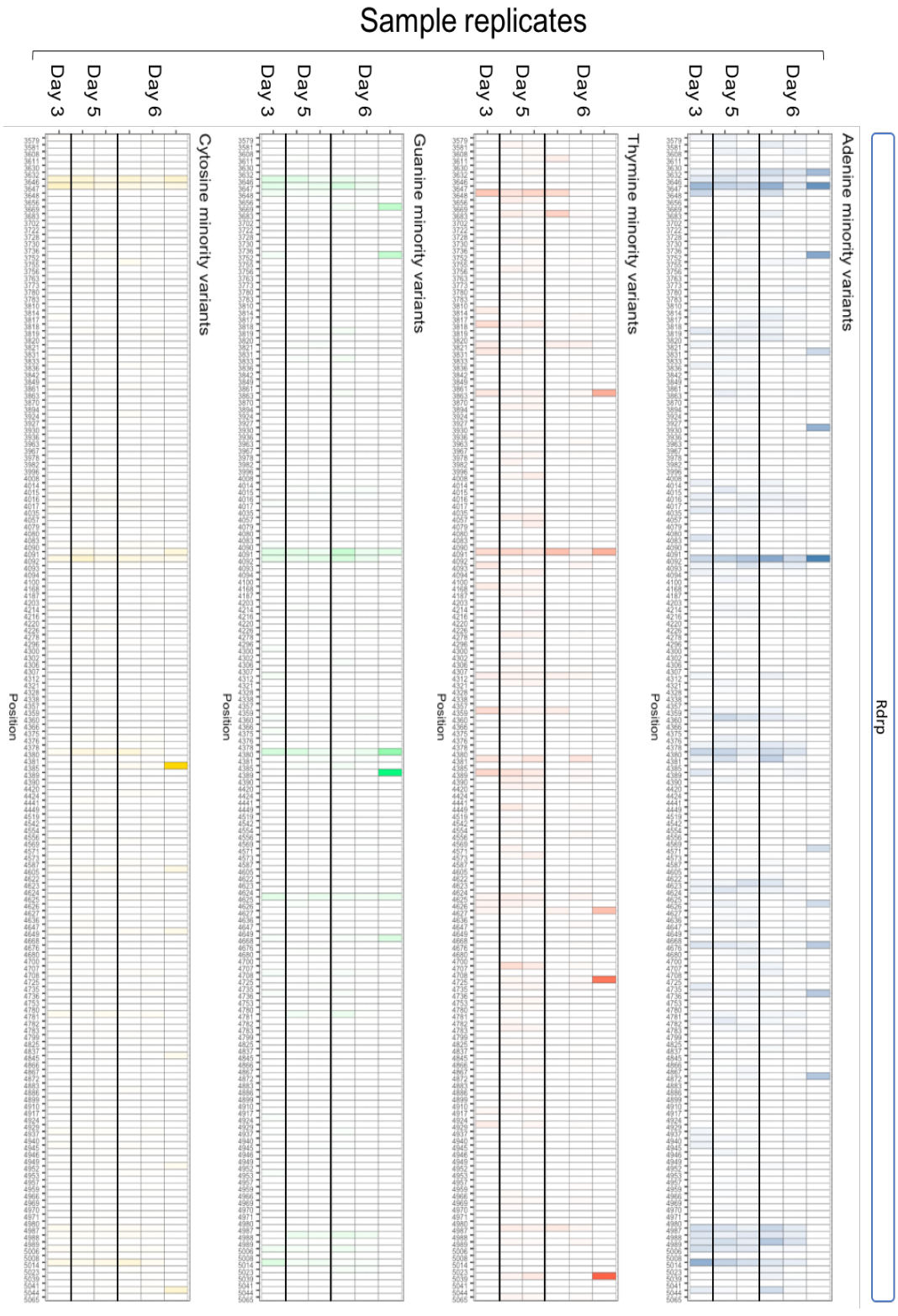


Figure 4.7 Minority variants located in the RNA-dependent RNA polymerase non-structural gene over the course of acute GII.4 infection in patient C.

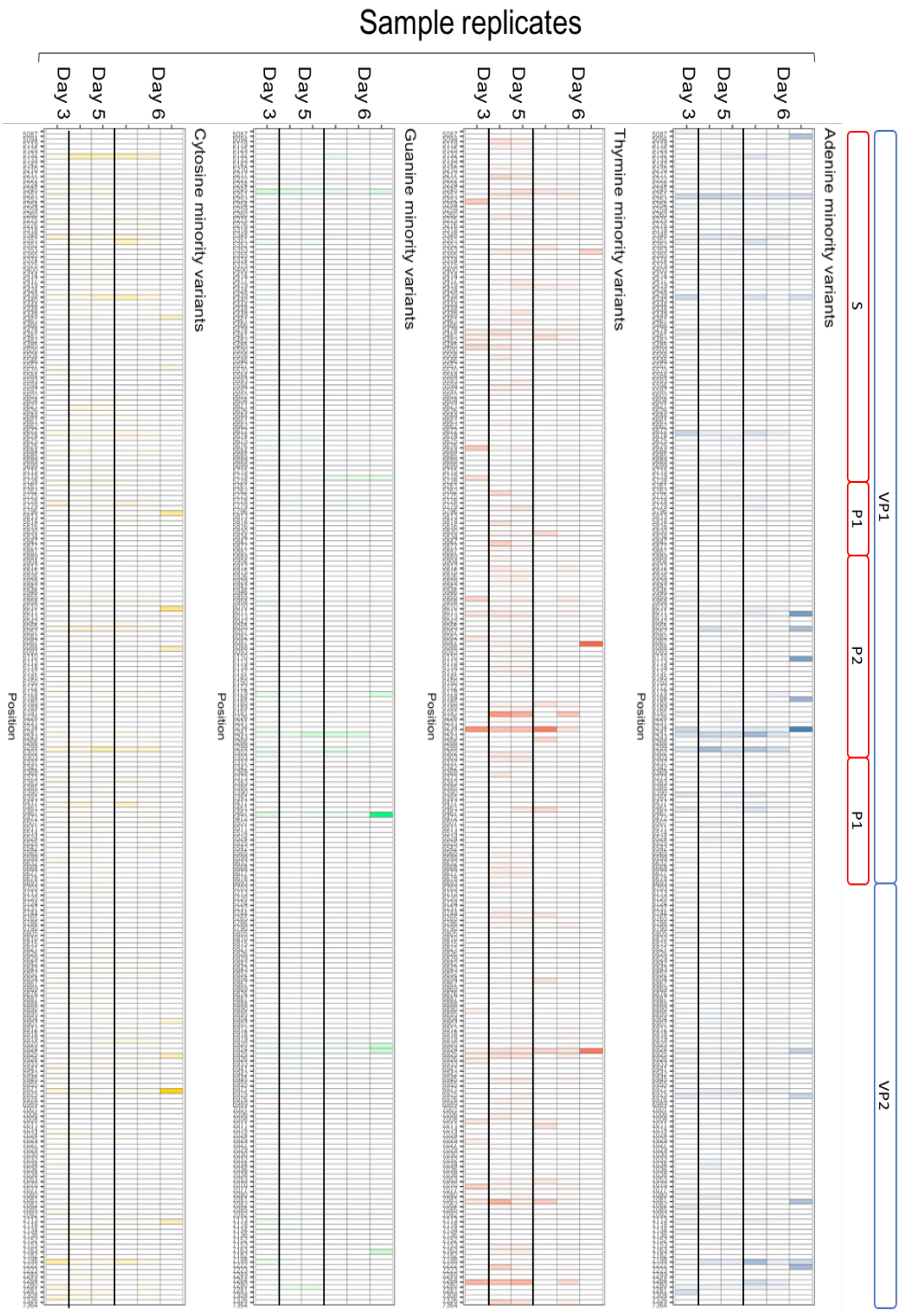


Figure 4.8 Minority variants located in the major (VP1) and minor (VP2) structural genes over the course of acute GII.4 infection in patient C.

Table 4.11 The number of persistent minority variants across the collection time points in each gene over the course of infection in patient B and C.

Gene		Frequency at gene	
		GII.4	GII.7
p48		1	5
NTPase		1	3
p22		2	4
Vpg		1	2
$3CL^{pro}$		3	1
RdRp		13	0
VP1	S	2	0
	P1	0	
	P2	3	
VP2		3	1

In agreement with analysis performed to determine which positions of the codon were diverse, most of the GII.7 minority variants were in ORF1 and one was found in ORF3, whilst most of the GII.4 minority variants were in ORF1, followed by ORF2 and then ORF3 (Table 4.11).

If a minority variant was found in all replicates from patient B at a single time point, there was a tendency for it to be detected at all time points, implying in GII.7 or chronic HuNoV infection there are underlying sets of variants that are maintained over short time periods (Figure 4.4). However, there were several exceptions, at positions 725, 2089, 2186 and 3128 where variant replacements have occurred (described previously) or the mutation can no longer be detected (Figure 4.4).

Although more persistent minority variants existed in the GII.4 population of patient C, there were also multiple examples of mutations present at a single time point only, which shows there are subpopulations that exist over the course of infection and others which move in and out of existence.

4.3.6 In chronic GII.7 illness, fewer minority variants were present as a larger fraction of the virus population in contrast to acute GII.4 infection

To better understand how each virus population was changing, the frequency of mutations at loci identified to contain persistent minority variants were measured as a fraction of the total number of base calls (Figure 4.9). Although there were fewer subpopulations in chronic GII.7 infection when compared to the GII.4 infection, those present occupied a much larger fraction of the virus population (Figure 4.9).

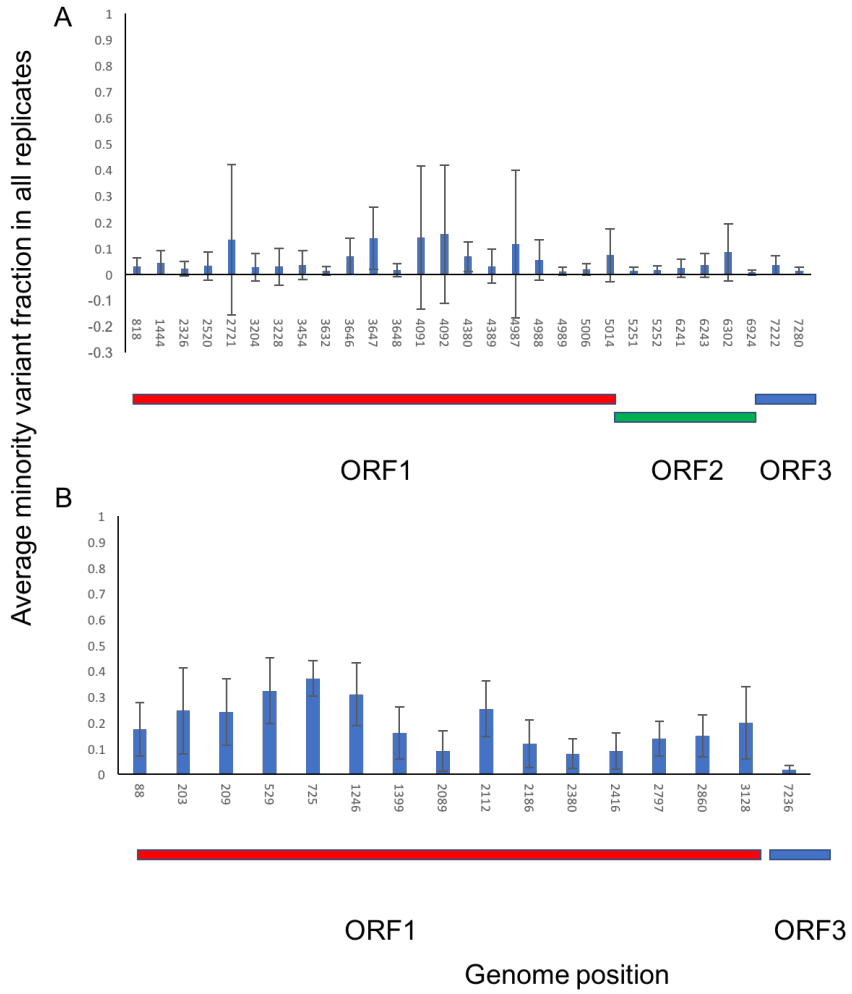


Figure 4.9 The mean proportion of persistent minority variants in acute GII.4 (A) and chronic GII.7 (B) infection.

4.3.7 Identification of true minority variants in replicate sequencing investigations can be used to create an inference threshold

The nucleic acid extracts obtained from the longitudinal samples of patients D, E and F were converted into library preparations once and, unlike patients A, B and C, not sequenced in triplicate. For patient's B and C, a true variant was defined as a mutation present in all three data sets, and therefore true and false variants could be compared to determine a base call frequency and total coverage threshold to discern mutants from errors.

In the replicate libraries originating from patient B and C it was possible to call true variants at a lower fraction of the total base calls if the coverage was higher (Figure 4.10). Greater separation was observed in a subset of variants present in the chronic GII.7 HuNoV infection in contrast to GII.4 acute infection (Figure 4.10). However, both data sets showed the same pattern and were combined to identify a threshold at which only true variants were called (Figure 4.11). In the combined data set there were 22079 mutations, of which 17583 (79.7%) were false positives and 4496 (20.3%) were true positives. Two separate criteria were applied to the base miscalls (true variants and sequence errors) present in a dataset of patient B and C combined, which were: ≥ 5 % of the total base calls and ≥ 200-fold coverage or ≥ 20 % of the total base calls and ≥ 50-fold. Independently, these two criteria incorporated regions of moderate and high coverage to select for 93 variable sites, of which 92 were classed as true positives (Figure 4.11). However, this represented a small fraction (2%) of the total true positives (Figure 4.11).

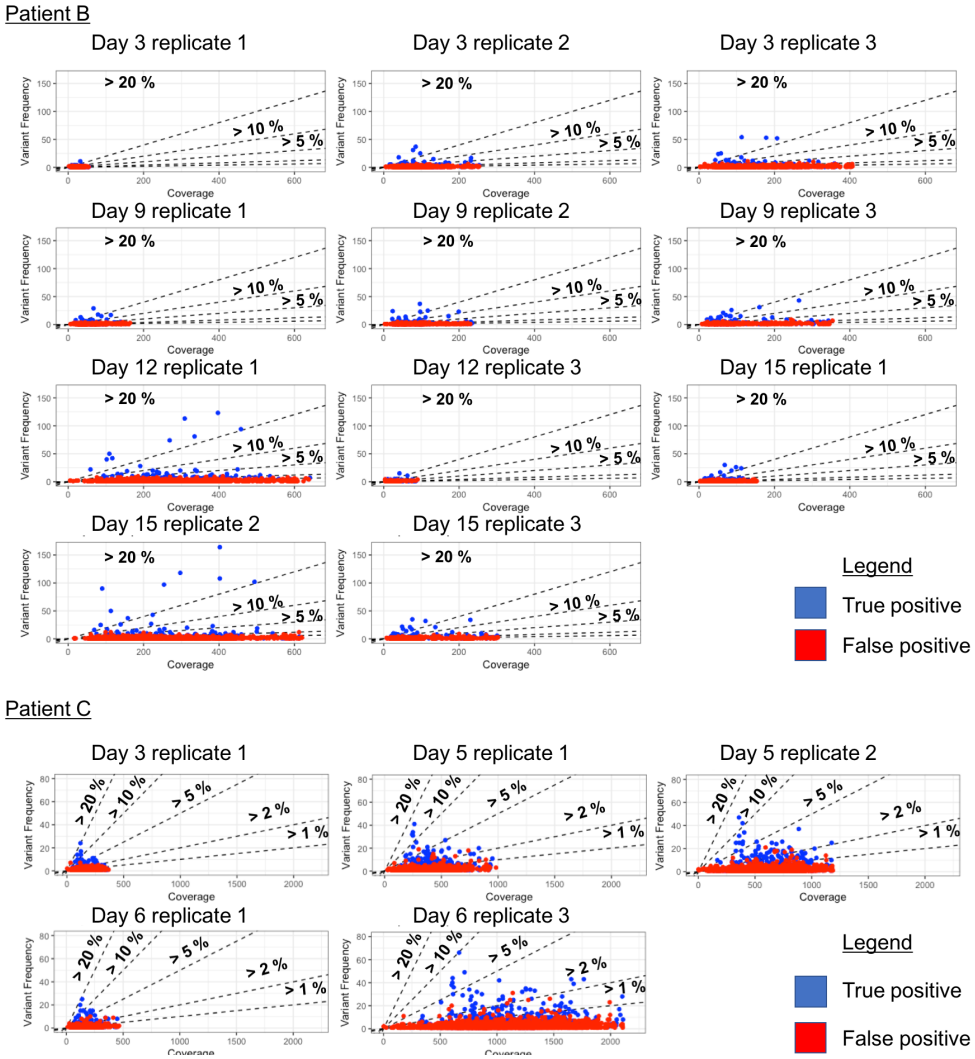


Figure 4.10 Comparison of variant frequency and total coverage of true and false minority variants.

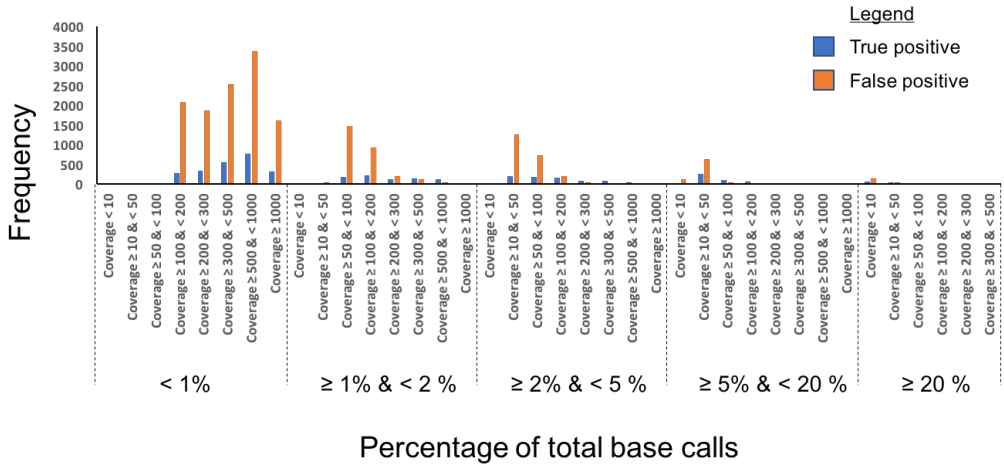


Figure 4.11 The frequency of true versus false positive variants called at different coverage thresholds.

4.3.8 Minority variants called in patient D, E and F

In total, 126 minority variants were found to be present across nine different time points in stool samples collected from patient's D, E and F (Figures 4.12 - 4.14). Most mutations were detected in patient E (76.9 %), followed by patient D (20.6%) and F (2.4%). Furthermore, the minority variant subpopulation was most diverse on day 0 in patient D, day 1 (the last time point) in patient E and day 0 in patient F. However, accurate base calling from the latter samples of patient F was likely restricted by the amount of sequence data available (Table 4.7).

In agreement with data from patient C, minority variants in patient D and E were present in every ORF, whilst those found in patient F were in ORF1 and 2 only (Table 4.12). Despite this similarity, the mutations detected in patient E were detected at a much higher frequency across the genome, and further, those in ORF1 were clustered (Table 4.12 and Figure 4.13).

In patient D, every minority variant identified by the criteria described was identified across the time points, but none replaced the wild type allele (Figure 4.12). However, in patient F, 2 of the 3 corresponding minority variants were not detected on day 2, and none were present on day 3 which is likely due to the limited amount of sequence data available (Table 4.7). In contrast, in patient E, the sample collected at the last time point showed allele replacement at various positions, in which the minority variants were identical to the consensus in the previous time points (e.g position 6189 - Figure 4.13). In addition, various positions had almost equivalent representation of two distinct variants (e.g position 2615 - Figure 4.13). Overall 52 strain replacements occurred, and just below a fifth of

these were in the P2 domain, whilst the remainder were present in ORF1

(Figure 4.13).

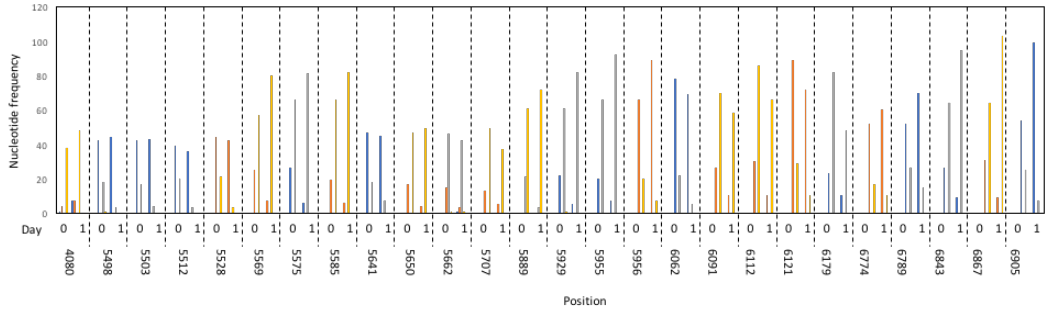


Figure 4.12 Allele frequencies at HuNoV GII.4 genomic loci identified to contain minority variants over the course of patient D infection.

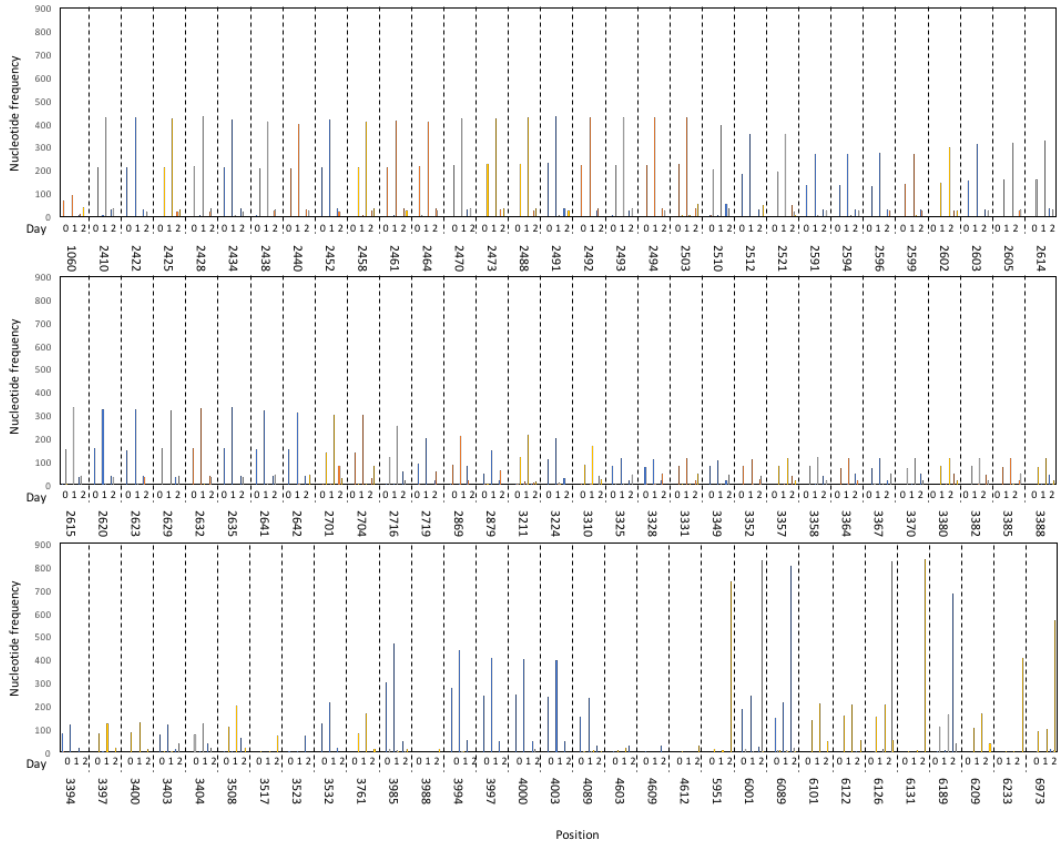


Figure 4.13 Allele frequencies at HuNoV GII.4 genomic loci identified to contain minority variants over the course of patient E infection.

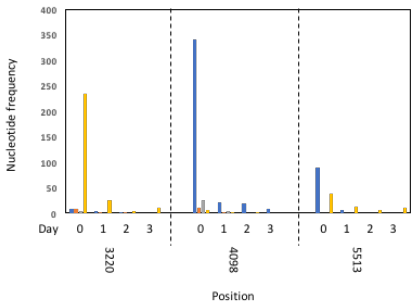


Figure 4.14 Allele frequencies at HuNoV GII.4 genomic loci identified to contain minority variants over the course of patient F infection.

Table 4.12 The number of minority variants across the collection time points in each gene over the course of infection in patient D, E and F.

Gene		Frequency at gene		
		Patient D	Patient E	Patient F
p48		0	0	0
NTPase		0	1	0
p22		0	33	0
Vpg		0	11	1
$3CL^{pro}$		0	26	0
RdRp		1	11	1
VP1	S	11	0	1
	P1	0	0	0
	P2	8	10	0
VP2		5	1	0

4.3.9 The sudden increase in diversity observed in patient E may have been caused by superinfection

The longitudinal samples obtained from patient E were aligned against the consensus sequence present on day 0. In the subsequent samples, genomic homogeneity was observed on day 1 (99.9% similarity), but at the last time point, collected on the same day, a sudden divergence in the virus population occurred (91.8% similarity). Interestingly, at the last time point, read coverage mapping to

the first two time points decreased in ORF1 but ORF2 and 3 were not affected (Appendix C).

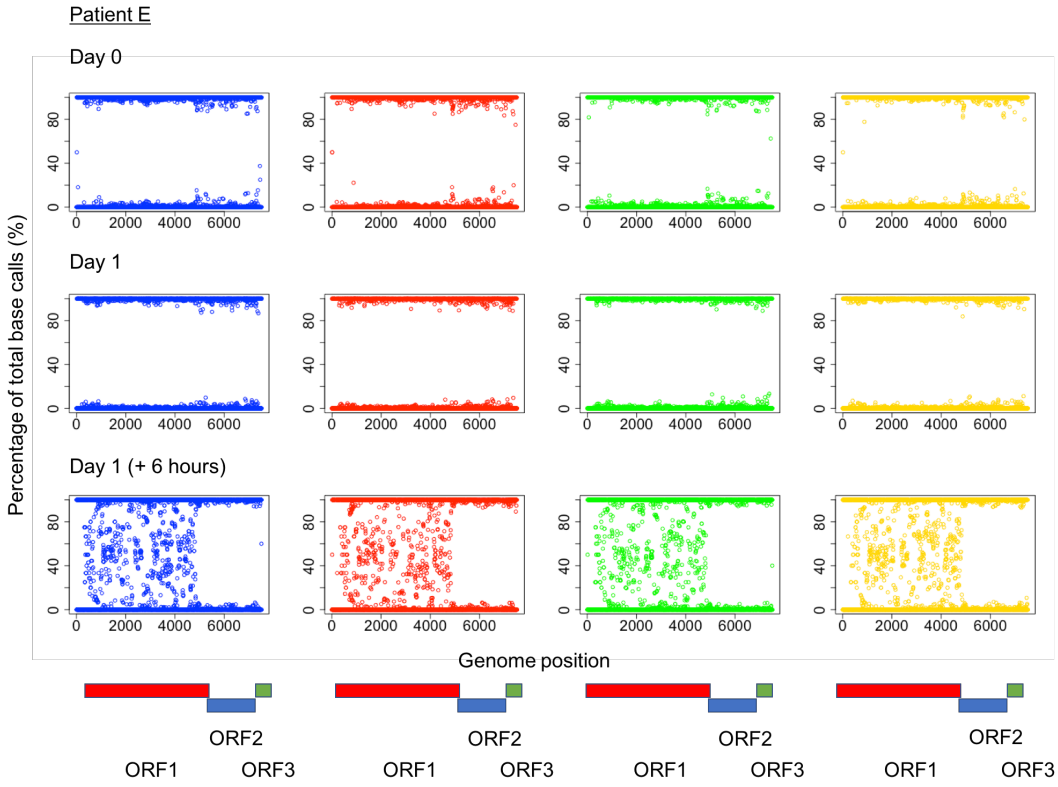


Figure 4.15 The proportions of base calls at genomic loci from patient E samples after alignment to the consensus sequence of day 0 (Blue=Adenine, Red=Thymine, Green=Guanine, Gold=Cytosine).

If the non-structural and structural genes were aligned separately, the ORF1 (90.1% similarity) was a more significant contributor to the increase in diversity than ORF2 and 3 (95.3%). To examine the change in diversity further, the proportions of alleles were measured across the genome in different samples (Figure 4.15). On day 0 and 1, predominant consensus alleles and minority alleles at low prevalence could be clearly discerned (Figure 4.15). However, on the latter time point of day 1 a mixed population of alleles was present in ORF1, whilst ORF2 and 3 remained homogenous (Figure 4.15).

On the last time point of patient E the consensus genome had changed markedly in contrast to the earlier two samples, and diversity appeared to be greater in the non-structural than structural genes, therefore *de novo* assembled contiguous sequences (> 500 bases) from each sample were mapped to each sample consensus sequence (Figure 4.16). On day 0 and day 1, there were two different populations a GII.16P/GII.4 recombinant and a GII.4. However, at day 1 (+ 6 hours) a contiguous sequence more like a complete GII.4 genome was detected, and recombinant sequences were still present but mapped to a very small proportion of the genome (Figure 4.16).

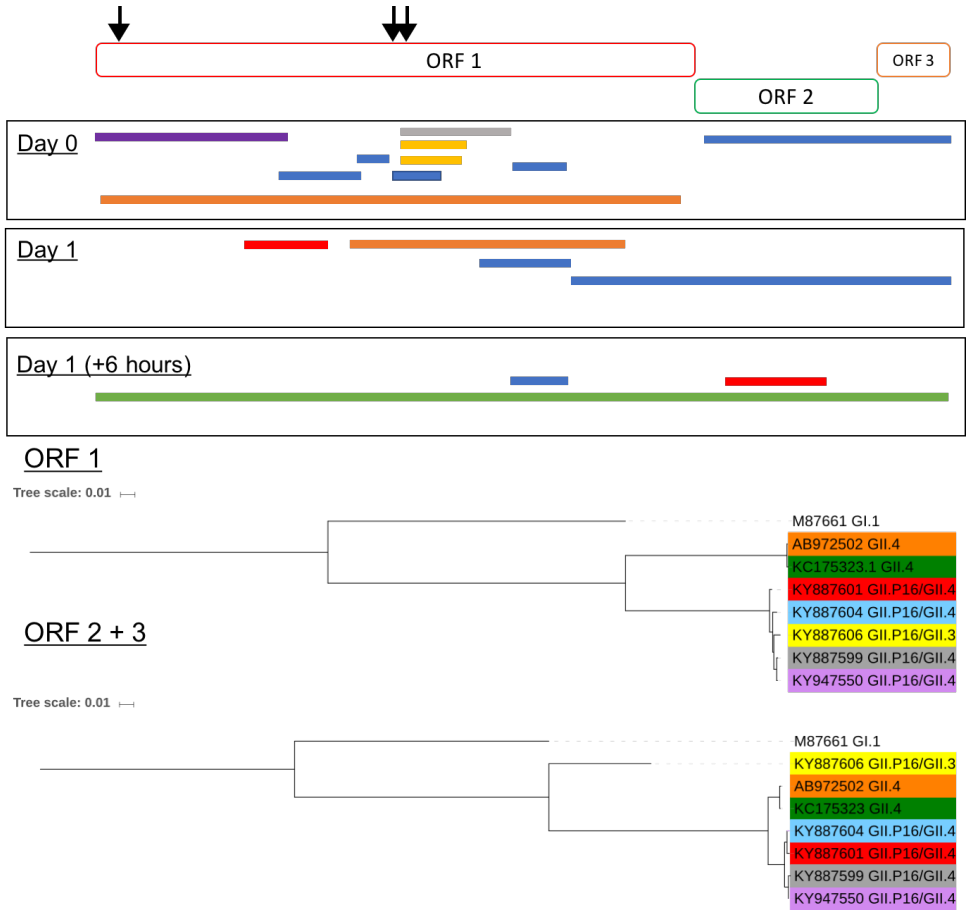


Figure 4.16 Mapping of de novo assembled contiguous sequences from patient E to the virus consensus sequences and phylogenetic analysis of the closest BLAST relative (The black arrows correspond to gaps in the consensus sequence equal to or below five after mapping).

4.3.10 Minority variants and consensus changes map to exposed sites in the P2 domain

Overall, four non-synonymous changes were detected in the HuNoV P2 subdomain. Two of these existed as minority variants in the early stages of patient C infection, whilst the remainder were actual consensus changes that occurred on the last time point of patient E infection.

Table 4.13 Mapped epitope locations on the GII.4 P domain in the literature. The colours correspond to the AA positions in the model below.

Publication	VP1 amino acid position																																										
	226	294	296	297	298	310	316	331	333	336	338	340	343	344	345	346	347	348	368	372	373	374	376	382	389	390	391	392	393	394	395	396	397	407	412	413	440	441	442	443	444	448	484	493
Epitope		A				N			B			C							A				C	B			D		D					E									N	
Allen et al, 2008																																												
Debbink et al, 2011																																												
Lindesmith et al, 2012																																												
Lindesmith et al, 2014																																												
Vicente, et al 2016																																												

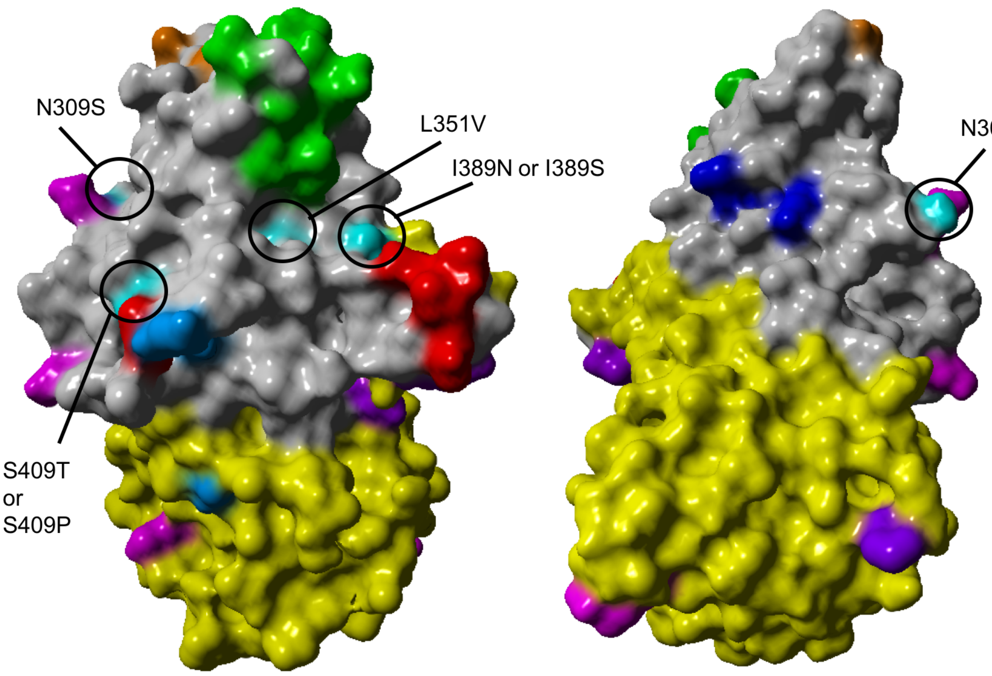


Figure 4.17 Epitope residue mapping of a GII.4 P domain mapped to X-ray crystallography data from the Protein Data Bank (PDB): 4OOS (Singh et al. 2015). Sites mapped from monoclonal antibody (mAb) and phage mapping studies in the literature (Table 4.13). The minority variant changes in patient C and consensus changes in patient E are labelled (cyan).

In patient C, two different non-synonymous mutations existed at each site over the course of infection, I389N or I389S and S409T or S409P. Position 389 is known to influence HBGA binding affinity (Tan and Jiang 2008), and is in a pocket which is adjacent to AAs that contribute to conformational epitope D and HBGA binding (Figure 4.17). Position 409 is also present in a pocket, but is closer to residues which form epitope E (Figure 4.17). The two consensus changes in patient E demonstrated a similar pattern, as residue 309 is next to a component of epitope N and 351 is in a pocket just below AAs which form epitope A (Figure 4.17).

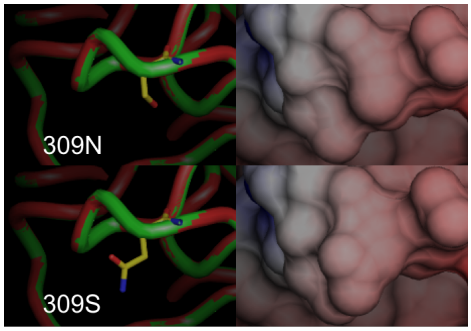


Figure 4.18 Protein modelling of the non-synonymous mutation at position 6001 (309) in patient E (Day 1 + 6 hours). Sequence mapped to X-ray crystallography data from the Protein Data Bank (PDB): 4OOS (Singh et al. 2015).

By modelling the non-synonymous mutations to an analogous GII.4 P domain structure it was possible to identify changes to the molecular surface (Figure 4.17 and 4.18). In patient E, L351V was buried within the antiparallel β-barrel of the P2 subdomain and did not appear to alter the surface structure (data not shown). On the other hand, N309S was present on an exposed loop and did change the molecular structure, but not the electrostatic surface charge (Figure 4.18). In patient C, none of the mutations altered the electrostatic surface charge, but

S409P, I389S and I389N did change the predicted shape of their respective cavities (Figure 4.19).

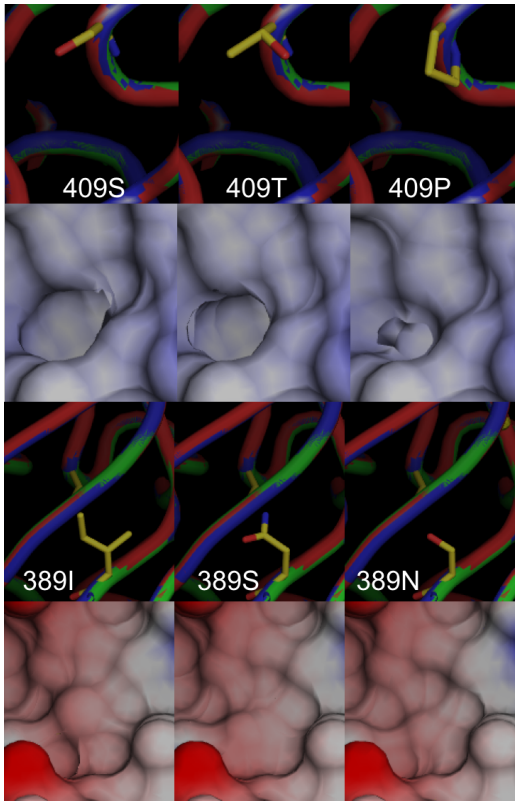


Figure 4.19 Protein modelling of the non-synonymous mutation at position 6243 (389) & 6302 (409) in patient C (Day 3,5 & 6). Sequence mapped to X-ray crystallography data from the Protein Data Bank (PDB): 4OOS (Singh et al. 2015).

4.3.11 Secretory Ab responses to a panel of time-ordered VLPs were highly variable between patients

To understand whether a relationship existed between the diversity of a HuNoV population and the host immune response at different stages of infection, a capture ELISA was developed to measure VLP specific copro-antibody responses in stool. To normalise the observations at different time points the specific HuNoV response was measured as a percentage of the total IgA. The average total IgA concentration in individuals with acute infection (1930±2338 μg/mL) was greater than in chronic

infection was (190±208 µg/mL). The proportion of HuNoV specific IgA/total IgA ranged from 0 to 7%.

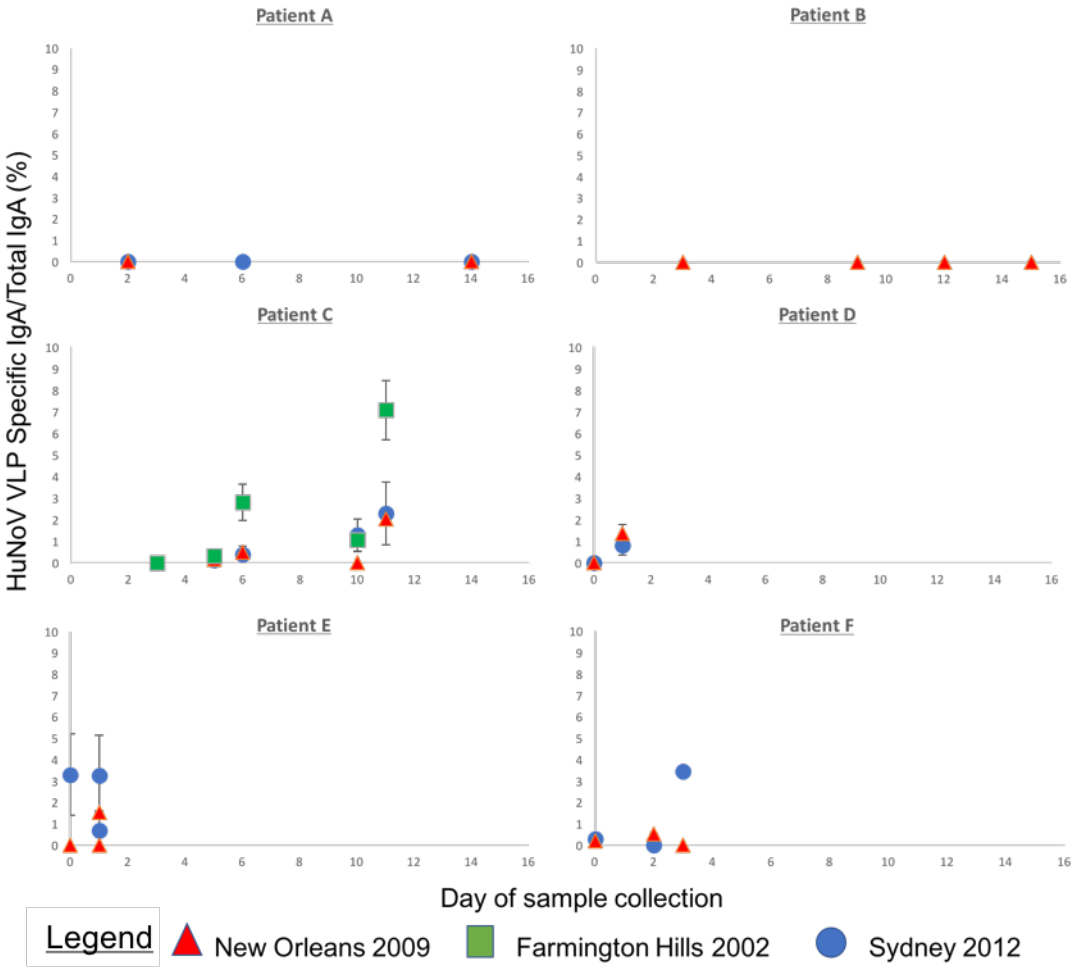


Figure 4.20 HuNoV GII.4 variant VLP specific copro-antibody responses over the course of infection.

The time ordered panel of GII.4 HuNoV VLPs consisted of Farmington Hills-2002 (FH), New Orleans-2009 (NO) and Sydney-2012 (SY). Stool samples from patient's A and C were analysed for reactivity to the whole panel, but samples from B, D, E and F were exposed to the Sydney and New Orleans strain only (Figure 4.20). An

insufficient amount of Farmington Hills VLP was available to assay all the patient samples, whilst a limited amount of stool sample from patient F on day 1 meant reactivity could not be tested against the panel.

GII.4 HuNoV-specific secretory IgA was not detected in patient's A or B at any time point (Figure 4.20). In patient's D and F, no response was detected to either the NO or SY VLPs on day 0, and in patient E no response was detectable to NO on day 0 or the first sample from day 1, but a response to NO was seen in the second sample from the same day (Figure 4.20). In patient C, no responses were detected until day 5 post infection, and from that point onwards an increase in the HuNoV specific response was observed at each time point with the response being predominant against FH, but also detectable and increasing against the NO and SY VLPs (Figure 4.20).

4.3.12 Correlation between GII.4 HuNoV-specifc copro-antibody and non-synonymous mutations during infection varied between patients

To further characterise interactions between the host immune response and infecting virus population, it was necessary to identify which GII.4 strain was the closest relative to the HuNoV shed by the patient, and any capsid antigenic similarities between the population and the VLP.

The GII.7 virus population shed by patient B had identical capsid sequences over the course of infection (Figure 4.21). In contrast the capsid sequences recovered from patients A, C, D, E and F clustered closest to the Sydney-2012 strain (Figure 4.21).

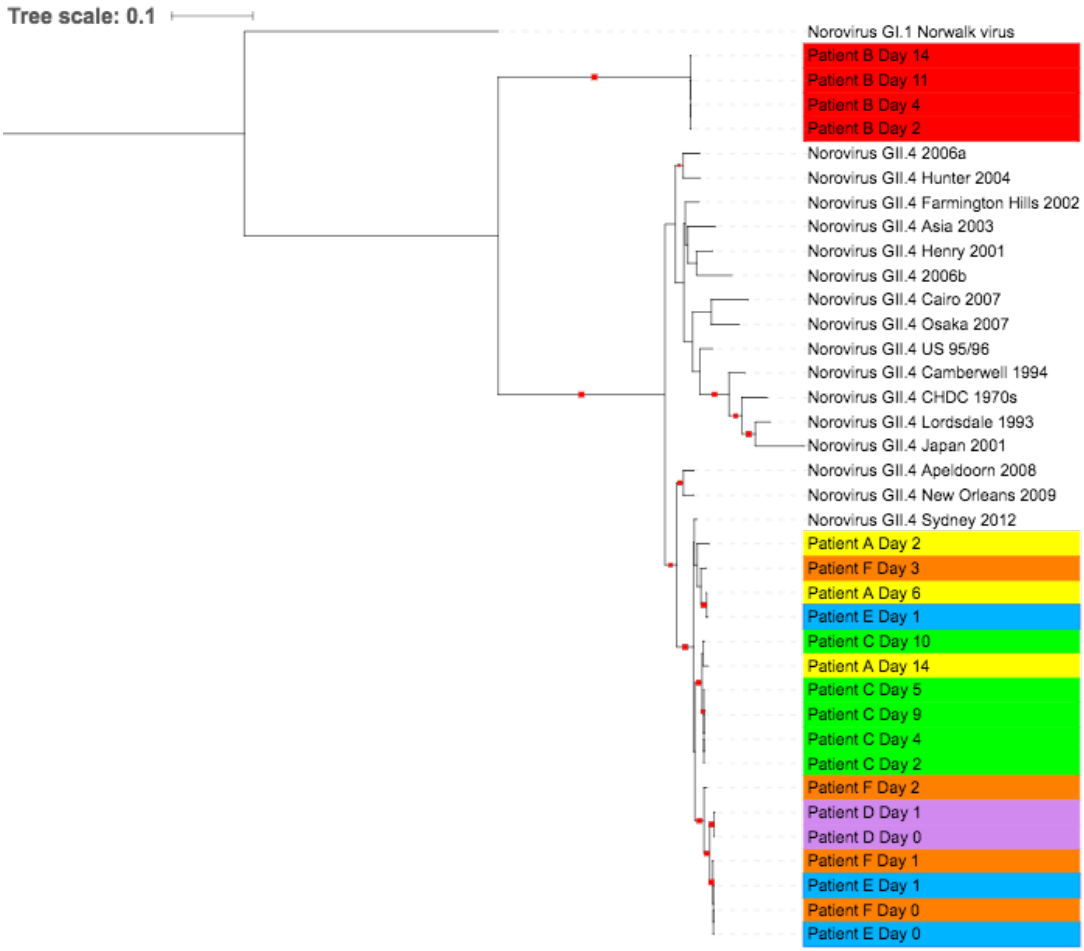


Figure 4.21 A rooted Maximum-Likelihood Tree of the capsid nucleic acid sequence for each patient sequence and GII.4 reference strains (boot strap values of nodes over 0.8 of 1000 bootstrap replicates are shown as red squares).

In agreement with the phylogenetic analysis, every virus population recovered was more closely related to the SY-VLP in the P2 domain than to NO or FH (Tables 4.14 and 4.15). In the infection of patient C, the P2 domain of the virus population was near identical to the SY-VLP in the P2 domain, with a single mismatch at AA position 329 (Table 4.14). However, if non-synonymous changes during infection were excluded, the virus populations present in patient D, E and F differed from patient C in the P2 domain at sites 323, 329 and 340, and likely reflects strain divergence over

time since the emergence and spread of the GII.4 Sydney strains (Tables 4.14 and 4.15).

Despite the virus identified at all-time points in patient C having the highest homology with the SY-VLP, the earliest and highest magnitude Ab response was directed towards FH, which was the most divergent. This was followed by the response to the homologous VLP, whilst the latest and lowest in magnitude was to the NO-VLP. In contrast, patient's D and F had detectable NoV-specific IgA to the homologous SY variant at days 1 or 2 post-date of onset, and patient E on the day of diarrhoea onset. The strains identified in these patients had a few mutations in the P2 domain, including a consistent A323T and a A340T, which has been identified in the literature as a putative epitope. In addition, further mutations along the P2 domain were mapped to putative epitopes in those patients where non-synonymous mutations were detected during infection: Patient E, 3/8 were present in sites known to be epitopes, and in patient F or D there were 2/5 and 1/2, respectively (Tables 4.14 and 4.15).

In patient D, the appearance of HuNoV-specific copro-antibody to NO or SY coincided with all the non-synonymous mutations observed during infection (Table 4.15). On the other hand, AA changes in patient E occurred on the last time point, despite the presence of SY-specific secretory IgA in all the stool samples collected (Table 4.15). Of the 5 non-synonymous mutations which were found in the virus population of patient F, a single epitope and non-epitope change was detected on day 3 when an increase in SY-specific copro-antibody was observed (Table 4.15).

Table 4.14 Conserved and variable AAs in the virus population of patient C sites identified by comparison with the VLP panel (Red=Epitope A, Blue=Epitope B, Green=Epitope C, Orange=Epitope D, Grey=Epitope E, Yellow=HBGA binding and CAb = Coproantibody).

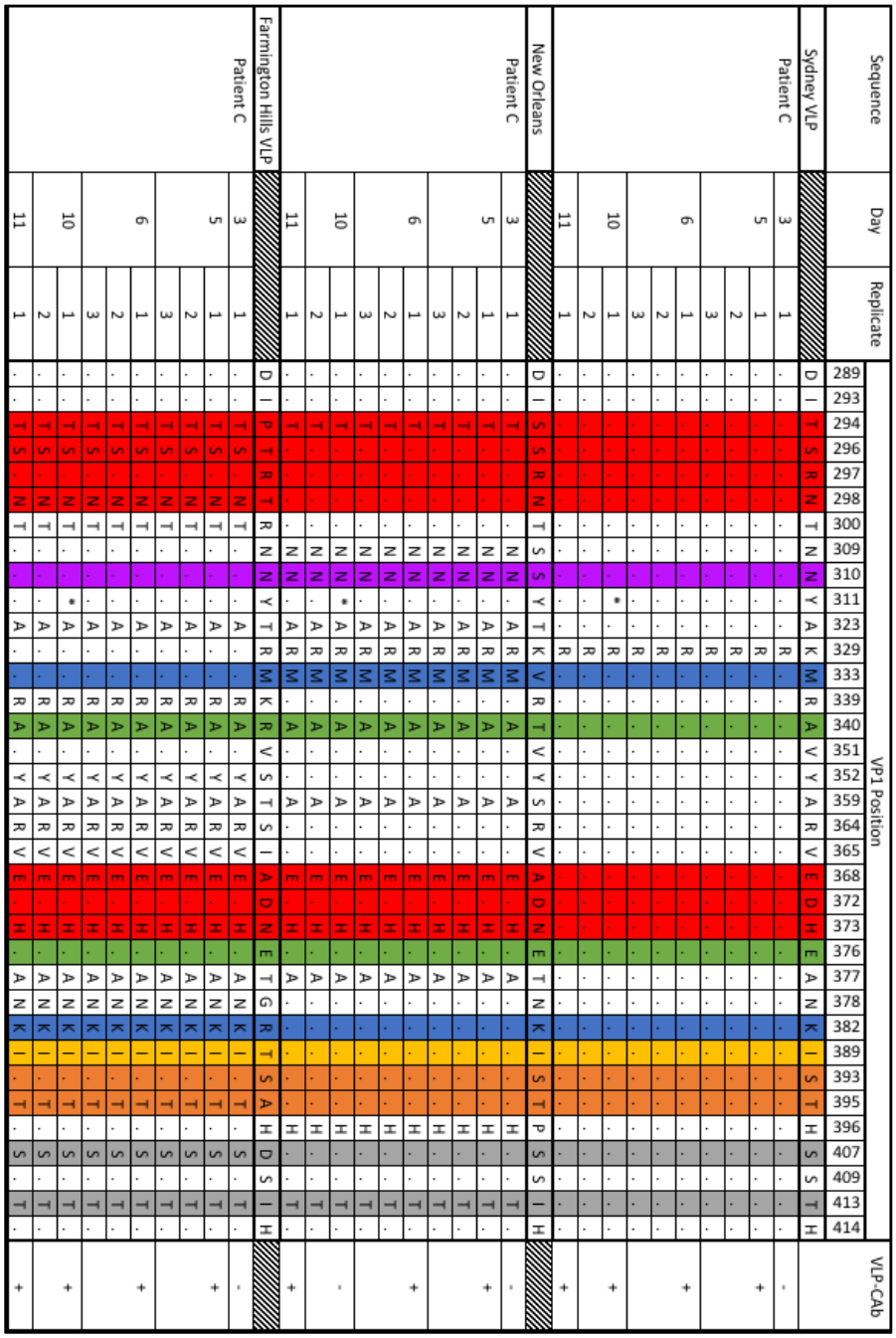
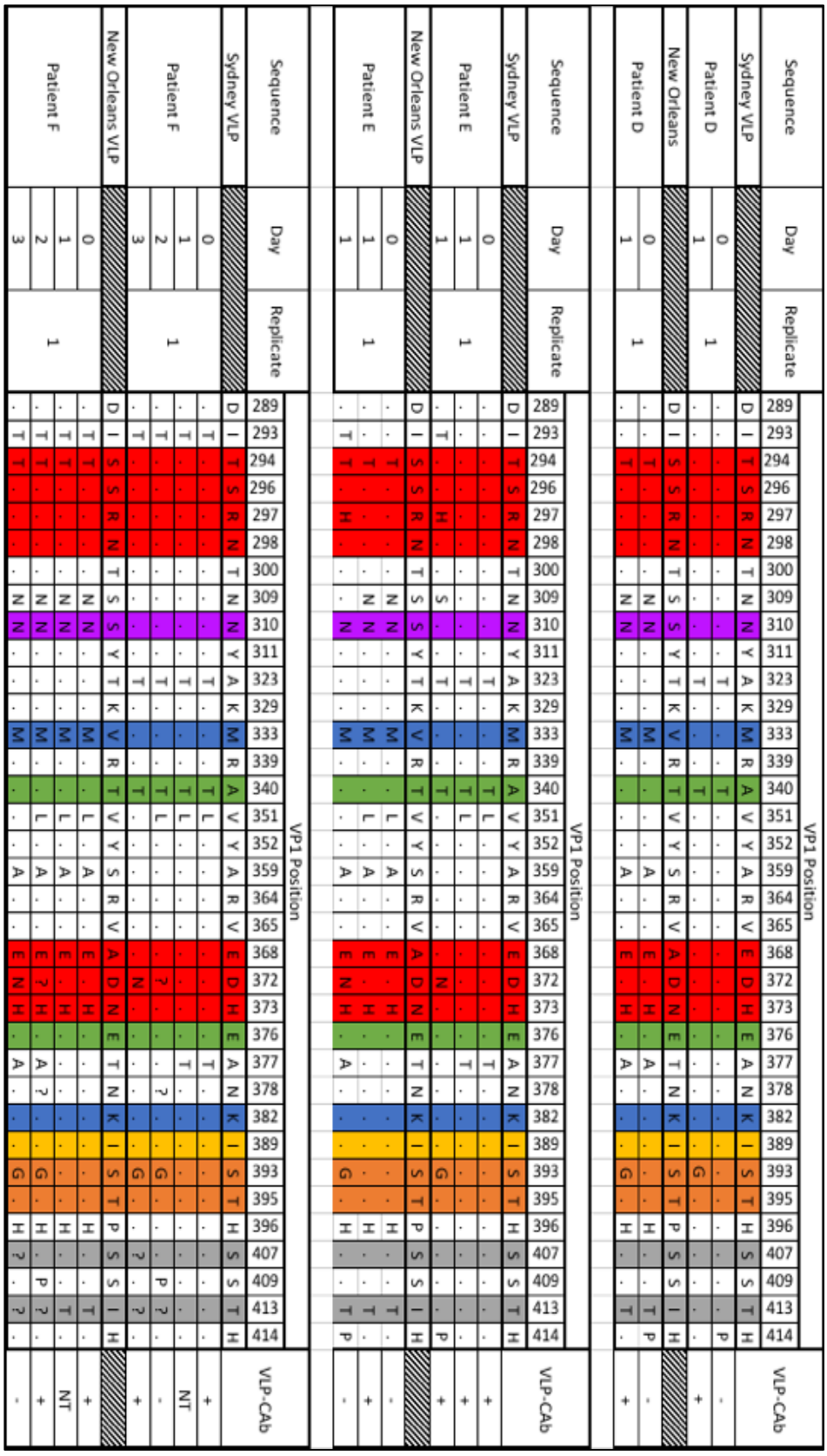


Table 4.15 Conserved and variable AAs in the virus population of patient D, E and F sites identified by comparison with the VLP panel (Red=Epitope A, Blue=Epitope B, Green=Epitope C, Orange=Epitope D, Grey=Epitope E, Yellow=HBGA binding, CAb = Coproantibody and NT=Not tested).



4.4 Discussion

4.4.1 HuNoV load in stool samples is critical for high genome coverage using MPS

In molecular biology, MPS technology provides a powerful tool to recover complete data from a nucleic acid extract, and then infer what is present, and whether it has undergone or is undergoing selective change. In this investigation, a total of 21 longitudinal samples from different patients underwent sequencing, and five GII.4 HuNoVs were identified in patient's A, C, D, E and F with acute illness, whilst the virus responsible for chronic infection in patient B was a GII.7.

Overall, library preparations were reproducible and PGM-MB capture meant near complete HuNoV genomes were recovered at the earlier stages of acute infection and throughout chronic infection. However, two library failures did occur, which did not provide any HuNoV reads or close to zero reads (Patient B Day 12 & Patient A Day 7). Prior to sample submission and pooling, each library was quality controlled by measuring the DNA concentration and targeting the GII HuNoV ORF1/2 region by qPCR (Kageyama et al. 2003). Therefore, failures could occur if the sample degrades, the ORF1/2 junction is overrepresented relative to the whole genome or the DNA concentration is not accurate and inadequate pooling occurs. Inter-replicate variability was present in those library preparations which were not considered to be failures, and could be attributed to PCR drift and the variability that exists between different sequencing runs (Verbist et al. 2015; Wagner et al. 1994). Although this variability existed, in individuals with acute infection there was a general trend of the HuNoV genome becoming more fragmented and decreasing

in mean genome coverage as shedding progressed over a week, which was not observed in chronic infection. This pattern correlated with reducing virus titres, and presumably progression of the infection into the convalescent stages, leading to higher coverage levels being inherently more difficult to attain. On the other hand, mean coverage in patient B likely remained consistent due to prolonged shedding of high viral loads which is well established in chronic HuNoV infection (Bok and Green 2012). An inability to recover whole genomes from non-longitudinal samples collected during acute infection has been described previously (Hasing et al. 2016).

4.4.2 In an acute GII.4 infection allele frequencies are stable and unlike those found in GII.7 chronic infection

To reliably identify genomic mutations in acute and chronic HuNoV infection, replicate library preparations from longitudinal patient stool samples underwent MPS. In a viral infection at a single time point, the virus with the greatest fitness will exist as the dominant allele, and therefore allele frequencies can be measured to detect strain replacement.

In agreement with a previous study, different allele frequencies in acute infection were stable in comparison to chronic infection (Bull et al. 2012). Overall, two synonymous mutations occurred in the ORF2 of the GII.4 over 11 days and a single synonymous mutation was present in the ORF1 of the GII.7 over 15 days. Hoffman and colleagues determined a nucleic acid substitution rate of 8.76 per 100 days by Sanger sequencing clonal isolates, which is in close agreement with that obtained (6.66 vs 8.76) (Hoffmann et al. 2012b). Moreover, Hoffman proposed the presence of positive selection by comparing sequences with published genomic data to

represent virus transmission in the population, however, this would be difficult to infer given the short time period over which samples were collected.

4.4.3 GII.7 in chronic infection and GII.4 in acute infection appear to be under selective pressure

HuNoV genomic loci that comprised heterogeneous base calls from the consensus sequence, and were present in all the replicates at a given time point, were separated by codon position to identify whether the virus populations in patient's A and C were changing under selective pressure. In patient C, an increase in the number of minority variants was observed from day 1 to 6 (replicate 1), respectively, whilst coverage remained similar, which could imply the mutant repertoire is expanding over time. Most of the minority variant diversity in patient C's GII.4 virus population led to non-synonymous mutations in ORF1, and a similar trend was observed in the GII.7 virus population of patient B. However, if the frequency of mutations at each codon position were normalized by gene size the difference between ORFs in patient C were not significantly different, which implies the selective pressure is equal across the non-structural and structural genes. But in patient B, the greatest frequency of mutations in the 1st codon position was in the ORF1, and when normalized by gene size the rate of mutation was significantly different to ORF2, but not ORF3.

In contrast to these data, a separate study identified purifying selection (846 dS/236 dN) in a cohort of immunocompromised individuals with chronic HuNoV infection (Kundu et al. 2013). However, the AA fixation rate has been shown to depend upon the extent of immunosuppression, and in immunocompetent

individuals the fixation rate is highest (0.13 AA/day) (Siebenga et al. 2008). This study agreed with Siebenga and colleagues, as patient C had more mutations per sample at all codon positions when compared to patient B (1236 vs 310). The low frequency of mutations in the major capsid protein of the GII.7 virus population would suggest that the immune system of patient B is not placing any selective pressure upon their virus. Interestingly, *in vitro* experiments performed by Bull and colleagues determined GII.7 viruses to have a faster RdRp incorporation rate than GII.4, but a lower mutation rate (Bull et al. 2010). Moreover, this study proposed the mutation and replication rate in GII.4 HuNoVs is optimal for adaption leading to global predominance, and hence why GII.7 tends to be detected in naïve paediatric populations (Bruggink et al. 2012; Bull et al. 2010; Puustinen et al. 2012; Qiao et al. 2017; Supadej et al. 2017). Despite these observations there are three limitations in making direct comparisons between patient B and C, and comparisons with other studies. Firstly, two different HuNoV genotypes are in separate environments and therefore it is difficult to elucidate whether virus population changes are induced by the host or an inherent property of the virus. Secondly, more replicates of adequate quality were available for patient B, which meant variant filtering will have been more stringent in relation to patient C. Thirdly, our samples were collected over a much shorter period when compared to other MPS studies of HuNoV in immunocompromised individuals (Bull et al. 2012; Kundu et al. 2013).

4.4.4 Minority variants in the GII.4 HuNoV population were spread over more genes at a lower frequency in contrast to GII.7

In patient B, minority variants were detected in ORF1 and 3, but none were present in the RdRp gene. However, in patient C minority variants were detected in all regions across the genome, but not in the P1 subdomain of ORF2. Although more minority variants were present across a broader selection of genes in GII.4 acute infection (patient C), those that were detected in GII.7 (patient B) chronic infection existed, on average, as a larger proportion of the virus population (0.05 vs 0.19).

In acute HuNoV infection, emergent minority variants were reported to largely exist at below 2.5% of the virus population, and over 10 days to change by 0.07% on average with a maximum variation of 2.6% (Bull et al. 2012). In this study, 11/29 of the minority variants were within this range, and the average was more than 10-fold higher (0.07 vs 5%). Over a much longer period of chronic infection (288 vs 15 days), Bull and colleagues demonstrated allele frequencies were much more variable (2.1 – 59.5%), which are similar proportions to those detected in patient B (1.8 – 37.1 %) (Bull et al. 2012). Furthermore, the 454 (Roche) instrument highlighted that the high allele frequencies comprised multiple separate haplotypes with common single nucleotide polymorphisms (Bull et al. 2012), however, this analysis was not possible as the HiSeq (Illumina) read lengths are much shorter and therefore difficult to reconstruct. But it was noted that in the acute infection of patient C there was a tendency for mutations to be present in all replicates, but not all time points when contrasted to the chronic infection of patient B.

4.4.5 An arbitrary cut-off for calling true minority variants can lead to data loss

The availability of replicate HuNoV MPS data sets, meant criteria could be designed to call minority variants in samples where replicates were not available. The criteria could identify true variants with 98.9% accuracy, but would also ignore 98% of the total true variants present if identified by using replicates instead. The criteria were applied to the longitudinal sequence data from patient E, D and F to identify 121 minority variants. In total 86 (71%) of the minority variants were present in at least one other sample from the same patient.

In comparison to patient C, patient D had a similar number of mutations (29 vs 25), and most were present in VP1. On the other hand, in patient F fewer minority variants were called, likely due to inadequate coverage (29 vs 3). In contrast, a sudden increase in consensus sequence diversity occurred in patient E on the last time point which coincided with an unusually high number of variants.

Several bioinformatics pipelines such as The Quality Assessment of Short Read Pipeline (QUASR) (Watson et al. 2013), Segminator II (Archer et al. 2012) and VirVarSeq (Verbist et al. 2015) have been developed and validated to call minority variants in virus populations, which can be difficult to distinguish from instrumental error. To validate the pipelines described, clinical samples or known standards have been analysed by one or two different sequencing chemistries, to measure the instrumental error rates. The instrumental mismatch error rates are in the range of 0.1-0.3 % for Roche 454 sequencing chemistry, 0.08-0.47 % for Illumina instruments and 0.12 % for the Ion Torrent PGM instrument (Archer et al. 2012; Gall et al. 2012; Van den Hoecke et al. 2015; Thys et al. 2015; Watson et al. 2013). Therefore, after

stringent filtering of the raw data, investigators have been able to apply these tools to their data and set thresholds above this range to accurately identify minority variants (Gall et al. 2013; Henn et al. 2012; Van den Hoecke et al. 2015).

It is important to note that prior to HiSeq (Illumina) MPS "spreading-of-signals" or "adapter leakage" can occur in the pool, where between 5-10% of reads are assigned to the incorrect sample index; this is due to low levels of free index primers, which are carried forward by AMPure XP beads (Beckman Coulter), and can be extended by DNA polymerase at the 3' end of a library even though index mismatches are present (Sinha et al. 2017). In this study, clinical samples were sequenced on the HiSeq platform, and therefore this is a potential limitation of calling minority variants. However, this would have been required to occur in an identical manner across several different time point samples, and in patient's B and C across two separate HiSeq (Illumina) MPS reactions.

4.4.6 A genotype mixture was present in early HuNoV infection of patient E

In the transition from the first two time points of Patient E's acute HuNoV infection to the last time point, the ORF1 consensus sequence changed from 99.9% similarity to 91.8%, which corresponded to a decrease in coverage and a mixed population of alleles in ORF1. Analysis of the *de novo* assembled contiguous sequences identified a mixed population of GII.4 HuNoV and a GII.P16/GII.4 recombinant. The GII.4 virus contiguous sequences mapped to the consensus genome at 68%, 32% and 100% coverage on day 0, 1 and 1 (+ 6 hours), whilst the recombinant covered 84%, 65% and 17%.

At the first-time point, a near complete GII.P16/GII.4 recombinant was assembled, alongside the ORF1 of a GII.4. Subsequently, on the last time point a near complete GII.4 HuNoV was assembled. In the literature, the ORF1/2 junction is well-established as a recombination hotspot (Bull et al. 2005; Bull et al. 2007). Further investigations with longer sequencing reads would be required to accurately reconstruct virus populations and discern whether a recombination event occurred, or whether competition existed between two separate HuNoV populations. It is possible the stool sample from patient E's last time point was mislabelled during collection, and was in fact from a separate patient, leading to the sudden increase in diversity observed. However, the presence of a near complete GII.4 ORF1 on day 0 in patient E is circumstantial evidence to suggest otherwise. Another possibility would be cross contamination during the library preparation process, which is difficult to exclude completely. Although a negative control was included in the GII ORF 1/2 junction qPCR, this was not subsequently included in the sequencing. MPS investigations comparing the concordance of HIV-1 sequences in plasma and peripheral blood mononuclear cells (PBMCs) have been able to exclude this possibility by phylogenetic branch separation between patient samples, but this would only be possible for patient B (Figure 4.21) (Lee et al. 2012).

Recombinants of GII.P16 have been detected in outbreaks and sporadic cases across Germany, Japan, South Korea, the United Kingdom and the United States (Cannon et al. 2017; Choi et al. 2017; Matsushima et al. 2016; Niendorf et al. 2017; Ruis et al. 2017). Furthermore, the GII.P16/GII.4 virus is considered a replacement of the predominant GII.17 in Asia (de Graaf et al. 2016). The HuNoV RdRp is a

determinant of virus transmissibility (Arias et al. 2016; Bull et al. 2010), moreover, four non-synonymous mutations in the GII.16P/GII.4 RdRp have been proposed to increase recombinant fitness leading to strain predominance (Ruis et al. 2017). This suggestion agrees with the early stages of patient E infection, where the GII.16P/GII.4 HuNoV appears to have outcompeted the GII.4, hence why the consensus on day 0 and 1 is recombinant-like, but on the last time point the GII.4 predominated and could coincide with differences in antigenicity. This finding highlights a major advantage of MPS over traditional sequencing, as the whole genome can be examined across a recombination hotspot and changes in allele frequency detected at high sensitivity.

4.4.7 Non-synonymous mutations in P2 subdomain of patient E coincided with increases in GII.4-specific secretory Ab

In total, 15 non-synonymous consensus changes occurred in the P2 domain of patient D, E and F's virus populations over the course of infection. Interestingly, S393G was found in all three patients, H414P was detected in patient D and E and mutations L351D, D372N, T377A were in common between patient E and F. Only a fraction (4/15) of these changes occurred upon the appearance of GII.4-HuNoV specific secretory antibody, whilst a further eight could have been induced after sustained selective pressure from the immune system. Overall, three of the AA changes (297, 372, 393) which could have coincided with the appearance of GII.4-specific secretory Ab, or from a sustained immune response, were present in previously described putative epitopes (sites A and D) (Allen et al. 2008; Debbink et al. 2012; Lindesmith et al. 2012a). The remaining sites which are not currently

described in the literature (293, 309, 351, 377, 414) could exist as uncharacterized epitopes with a role in HuNoV escape from herd immunity. However, the sites described should be interpreted with caution as, at low levels of coverage which were observed, toward the convalescent stages of acute HuNoV infection erroneous base calls can be determined as the virus consensus sequence.

Two of the consensus changes described in patient E were present at an earlier stage of infection as minority variants (309 and 351). Whilst in patient C, two minority variants were found (389 and 409) that had the potential to change the P2 subdomain surface structure, but were not detected on day 10 and 11 and could be a result of under sampling. The mutations present at positions 309, 389 and 409 were predicted to alter the P2 subdomain surface structure after protein mapping. Moreover, the AAs present at 309, 389 and 409 were near epitopes N, D and E respectively, and thus provided further evidence for these regions to have a role in GII.4 HuNoV escape from herd immunity (Allen et al. 2008; Debbink et al. 2012; Lindesmith et al. 2014; Lindesmith et al. 2012a).

4.4.8 A dominant copro-antibody was induced to an earlier GII.4 strain after exposure to a contemporary GII.4 strain

In patient C, the virus population was near identical to the SY VLP in the P2 domain with one AA difference at position 329. However, a stronger GII.4 secretory-Ab was induced toward the FH VLP on day 6 and 11, which was highly dissimilar (11/20 differences at known epitopes).

Two concepts exist that could explain the secretory Ab measurements observed in patient C, which are known as original antigenic sin (OAS) (Francis 1960) and

Antigenic Seniority (Lessler et al. 2012). OAS is the theory that the naïve host immune system is imprinted by the first antigen it encounters, and thereafter the highest Ab affinity responses will be induced to the first antigen when similar antigens are encountered (Francis 1960). On the other hand, Antigenic Seniority proposes that every similar antigen encountered form a hierarchy over time, and the Ab affinity for each is boosted by encountering similar antigens (Lessler et al. 2012). Antigenic Seniority was proposed to have occurred after the administration an intramuscular HuNoV VLP vaccine (consensus VLP of contemporary strains), where cohorts over 18 years old mounted a persistent response to GII.4-1997 (Lindesmith et al. 2015). However, Tamminen and colleagues demonstrated that exposure to GII.4, followed by GII.12 (after 14 weeks), did not affect GII.12 IgG titres or avidity in mice (Tamminen et al. 2013). Contrasted to this study, GII.4 and GII.12 are more distantly related than GII.4 FH and GII.4 SY, and therefore it is possible that the GII.12 was dissimilar enough to evoke a separate immune response. In the surrogate neutralisation assays, ten times more FH VLP was applied to the PGM receptor, which cannot be excluded as an explanation. But both concentrations (2 and 20 μg) would have been expected to saturate the PGM receptors and the excess would be washed away.

It must be noted that differences in phylogenetic clustering (Figure 4.21) of capsid sequences from the same patient at different time points could be due to a superinfection, an allele replacement within the virus population or a low-quality library preparation (decrease in read coverage).

4.4.9 Despite the presence of Sydney strain specific coproantibody patients D, E and F were infected by a novel HuNoV strain

Patients D, E and F were infected by a HuNoV strain which was phylogenetically closely related to the Sydney 2012 strain (Figure 4.21), even though SY-VLP specific coproantibody was present and most likely originated from a previous infection. Therefore, the three separate virus populations in these individuals were sufficiently different in the P2 domain to escape immunological memory and induce GE. In the P2 domain of each virus population, two conserved non-synonymous mutations existed as dominant alleles over the course of infection, A323T and A340T. The latter has been described as a component of a putative epitope on the virus surface, whilst AA position 323 has not been reported and therefore could potentially be a site of antigenic importance (Lindesmith et al. 2012a).

Understanding how the diversity of a HuNoV population changes in an immunosuppressed/immunodeficient host in comparison to an immunocompetent host can lead to the identification of regions or putative epitopes that may be under direct immune pressure. In the absence of a suitable and readily available cell culture or animal model, this approach may provide a way forward for identifying epitopes, and changes to the virus capsid that drive HuNoV persistence.

5 Whole genome sequencing as a tool for tracking transmission during outbreaks

5.1 Introduction

5.1.1 The importance of tracking pathogen transmission in public health

In public health, it is important to identify pathogens present in the population and their transmission routes, which can lead onto effective intervention strategies and disease control. Epidemiological investigations can infer the source of infection. However, high resolution pathogen genomic characterisation coupled with epidemiological data can increase the accuracy of the investigations and distinguish transmission events from multiple seeding events to improve intervention strategies.

5.1.2 Sanger sequencing as a tool for tracking virus outbreak transmission

The earliest developed tool for characterizing the genomes of viruses in point source and propagated outbreaks is Sanger sequencing, which is still commonly used to sequence short variable regions with high phylogenetic resolving power (Sanger et al. 1977). Sanger sequencing technology can be applied directly to amplicons or indirectly by cloning extracted DNA sequences into a plasmid vector. This technique can determine the complete genome sequence of the predominant virus in a population, but requires bridging of multiple amplicons.

Partial sequencing of an off-season Dengue outbreak in Kerala was performed by Anoop and colleagues, which led to the conclusion that the virus was phylogenetically very like another Dengue virus previously detected in a separate

part of India (Anoop et al. 2012). More recently, partial and complete genome sequencing aided transmission mapping of an Ebola outbreak in Guinea (Baize et al. 2014). In a similar manner, Sanger sequencing of complete Measles virus (MV) genomes was used retrospectively to characterize an outbreak of MV at the 2010 Olympic Winter Games (Gardy et al. 2015). MV was found to undergo strong purifying selection, and often the virus transmitted from person-to-person was 100% identical, but enough diversity was present over a short time period to ascertain spatiotemporal clusters (Gardy et al. 2015).

5.1.3 MPS as a tool for tracking outbreak virus transmission

In contrast to Sanger sequencing technology, MPS is not primer mediated and can provide complete genomes at much greater coverage depth. The depth of sequence data obtained can further inform on short regions of high variability (Penedos et al. 2015), and detect minority variant transmission from donor to recipient to improve the resolution of phylogenetic relationships (genetic bottleneck) (Montoya et al. 2016; Redd et al. 2012). Furthermore, the MinIon (Oxford Nanopore Technologies) platform was used in the recent Ebola virus outbreak to recover complete genome sequence data and inform on the geographical spread of variants within and between African countries (Quick et al. 2016). In a similar manner, two other MPS platforms, the 454 (Roche) and the Ion torrent (Fisher Scientific) instruments, have been applied to Zika virus to define the introduction of different strains across South America (Faria et al. 2016).

5.1.4 Sanger sequencing and MPS in HuNoV outbreaks

In the literature, partial Sanger sequencing of the HuNoV genome is an established epidemiological tool to assign genotypes to sporadic cases of illness within the population or causative agents in outbreaks, and further transmission. The earliest investigation to characterize transmission in separate local HuNoV outbreaks amplified and sequenced 3,225 nucleotides from the 3' terminal, and proposed person-to-person transmission to coincide with an accumulation of mutations as opposed to infection from a common source (Dingle and Norovirus Infection Control in Oxfordshire Communities 2004). On the other hand, Xerry and colleagues partially sequenced a shorter region which could be equally informative, the HuNoV P2 domain or hypervariable region (600 bases), in several studies (Xerry et al. 2010; Xerry et al. 2009; Xerry et al. 2008). These investigations led to the definition of a common source outbreak as 100% identical between susceptible individuals. However, in a small proportion of individuals (10%) 1-2 nucleotide changes can occur 3 weeks after the onset of shedding, which would reduce sensitivity if the mutated virus was transmitted further (Sukhrie et al. 2013).

Despite the establishment of MPS platforms in molecular biology, only a handful of studies have applied this technology to HuNoV transmission. Bull and colleagues partially sequenced the structural genes of a HuNoV recombinant transmission event between three immunocompetent individuals, with the 454 (Roche) instrument (Bull et al. 2012). The major alleles present in both recipients, were found as minor variants in the donor (< 0.01%) (Bull et al. 2012). A separate study, using the same instrument, also characterized a genetic bottleneck transmission

event between two immunocompromised patients by complete genome sequencing of HuNoV (Kundu et al. 2013). More recently, the complete HuNoV genome of a recombinant was characterized, with the Ion Torrent instrument (FisherScientific), in a subset of individuals infected in a presumed point source outbreak which occurred in the 1970s (Johnson et al. 2017).

5.1.4.1 Point source versus secondary transmission

A point source outbreak occurs when individuals encounter HuNoV indirectly through a contaminated fomite (i.e foodborne, waterborne or food handler transmission). Further, if a point source outbreak is induced by sewage contamination the genotypes detected will tend to be more heterogeneous, in contrast to food handler transmission (Le Guyader et al. 2006). Le Guyader and colleagues elegantly traced an international, mixed genotype, HuNoV outbreak to an oyster farm, contaminated by sewage, through epidemiological analysis and partial sequencing of region C and the P2 domain (Le Guyader et al. 2006). Moreover, partial sequencing of region A and C linked frozen strawberries to a multistate HuNoV outbreak in Germany, likely to have been caused by sewage contamination (Hohne et al. 2015). Partial sequencing of region C and the P2 domain recently indicated a high prevalence of asymptomatic healthcare workers or food handlers existed in point source outbreaks (Sabria et al. 2016).

Most HuNoV infections are transmitted through person-to-person transmission (i.e nosocomial transmission). Transmission of HuNoV between patients has been reported rarely, and is more difficult to characterize than common source outbreaks. A mutation which has occurred in one individual must be conserved in a

latter infection for circumstantial evidence of person-to-person transmission (Kundu et al. 2013; Sukhrie et al. 2010).

5.2 Materials and methods

5.2.1 Clinical sample preparation, enrichment, library preparation and sequencing

Faecal suspensions were prepared from 38 archived samples from various outbreaks gifted from Public Health England (PHE), for which epidemiological data had been obtained retrospectively. The stool suspension supernatant enrichment methodology, RNA extraction protocol and post-extraction processes (including library preparation) were identical to that described in Chapter 4, whilst MPS was performed in an identical manner to Chapter 3. In total 38 samples were prepared, however, sufficient MPS data to infer the P2 domain consensus sequence was only obtained in 25/38 library preparations (Tables 5.1 - 5.6).

Table 5.1 Stool sample details from outbreak A (* = RNA extract not positive for GII but was present after repeated library preparation).

Outbreak	Environment	Sample	Genotype	GII HuNoV Ct value
A	Restaurant	1	GII.20	23.16
		2		24.86
		3		27.88
		4		NA*
		5		25.33
		6		24.95

Table 5.2 Stool sample details from outbreak B.

Outbreak	Environment	Sample	Genotype	GII HuNoV Ct value
B	Hospital	1	GII.4	18.75
		2		21.99
		3		24.86
		4		24.38

Table 5.3 Stool sample details from outbreak C.

Outbreak	Environment	Sample	Genotype	GII HuNoV Ct value
C	Pub	1	GII.Pe/GII.4	21.74
		2		26.88
		3	GII.4	32.41
		4		22.54

Table 5.4 Stool samples details from outbreak D.

Outbreak	Environment	Sample	Genotype	GII HuNoV titre/qPCR reaction
D	Hospital	1	GII.4	29.00
		2		31.27
		3		30.24
		4		28.37

Table 5.5 Stool sample details from outbreak E.

Outbreak	Environment	Sample	Genotype	GII HuNoV titre/qPCR reaction
E	Hospital	1	GII.4	37.11
		2		14.09
		3		24.89

Table 5.6 Stool sample details from outbreak F.

Outbreak	Environment	Sample	Genotype	GII HuNoV titre/qPCR reaction
F	Hospital	1	GII.Pe/GII.4	20.64
		2		33.95
		3		21.15
		4		26.53

5.2.2 Data analysis

5.2.2.1 Bioinformatic analysis

The HiSeq (Illumina) data was analysed using the pipeline described in Chapter 3, whilst the identification of true minority variants was achieved by applying the criteria mentioned in Chapter 4. Multiple alignment analysis was performed in

Geneious 8.0.5 (Kearse et al. 2012) – www.geneious.com, with the MUSCLE algorithm (Edgar 2004) under default settings.

5.2.2.2 Phylogenetic analysis

Phylogenetic relationships were inferred as described in Chapter 4.

5.3 Results

5.3.1 A higher HuNoV virus titre correlated with a greater depth of read coverage

HiSeq (Illumina) reads from outbreaks A, B, C, D, E and F were assembled into contiguous sequences *de novo*, and a local nucleotide alignment showed the aetiological agents were all genogroup II HuNoVs, in agreement with the partial sequence data obtained upon sample collection (data not shown). In half of the outbreaks, which were all health care environments, a non-recombinant GII.4 HuNoV was present in all the stool samples collected (Tables 5.2, 5.4 and 5.5). In the remaining environments, the GII.Pe/GII.4 recombinant was detected alongside a GII.4 in a pub (Table 5.3), whilst the same recombinant was detected solely in a hospital (Table 5.6), and a GII.20 was assembled from the stool samples collected in a restaurant (Table 5.1). Consensus sequences were derived individually after mapping each set of reads to their respective reference sequence.

Complete or near complete HuNoV genomes were recovered from most stool samples (22/25) via PGM-MB enrichment, and none were classed as library preparation failures. Two of the three samples for which a fragmented GII HuNoV genome was recovered belonged to the same outbreak (Outbreak E), but did not

appear to be associated with the genotype present (Table 5.11). In total, six samples across all outbreaks had a median coverage below ten, and half of these were from the outbreak E (Tables 5.7 – 5.12). Over half of the samples (56%) had a median coverage above 100 (Tables 5.7 – 5.12), and in two of the samples the lower quartile range was greater than the upper quartile range (Figure 5.1).

Table 5.7 Summary of GII.20 reads recovered from stool samples obtained in outbreak A (IQR=Interquartile range).

Outbreak	Sample	Coverage		Percentage of the genome covered (%)
		Median	IQR	
A	1	520	617	99.67
	2	8	11	95.33
	3	331	479	99.57
	4	218	252	99.83
	5	172	219	99.73
	6	49	54	98.54

Table 5.8 Summary of GII.4 reads recovered from stool samples obtained in outbreak B (IQR=Interquartile range).

Outbreak	Sample	Coverage		Percentage of the genome covered (%)
		Median	IQR	
B	1	74	69	99.44
	2	1600	760	99.88
	3	203	146	99.74
	4	201	108	99.56

Table 5.9 Summary of GII.Pe/GII.4 and GII.4 reads recovered from stool samples obtained in outbreak C (IQR=Interquartile range).

Outbreak	Sample	Coverage		Percentage of the genome covered (%)
		Median	IQR	
C	1	12	13	95.75
	2	103	79	99.54
	3	7	7	94.12
	4	1128	744	100.0

Table 5.10 Summary of GII.4 reads recovered from stool samples obtained in outbreak D (IQR=Interquartile range).

Outbreak	Sample	Coverage		Percentage of the genome covered (%)
		Median	IQR	
D	1	4	5	85.06
	2	16	14	98.19
	3	179	191	99.91
	4	349	225	100.0

Table 5.11 Summary of GII.Pe/GII.4 reads recovered from stool samples obtained in outbreak E (IQR=Interquartile range).

Outbreak	Sample	Coverage		Percentage of the genome covered (%)
		Median	IQR	
E	1	3	3	83.85
	2	2	2	54.25
	3	4	5	90.24

Table 5.12 Summary of GII.4 reads recovered from stool samples obtained in outbreak F (IQR=Interquartile range).

Outbreak	Sample	Coverage		Percentage of the genome covered (%)
		Median	IQR	
F	1	3008	1693	99.99
	2	13	16	95.70
	3	3251	1908	100.0
	4	260	219	99.77

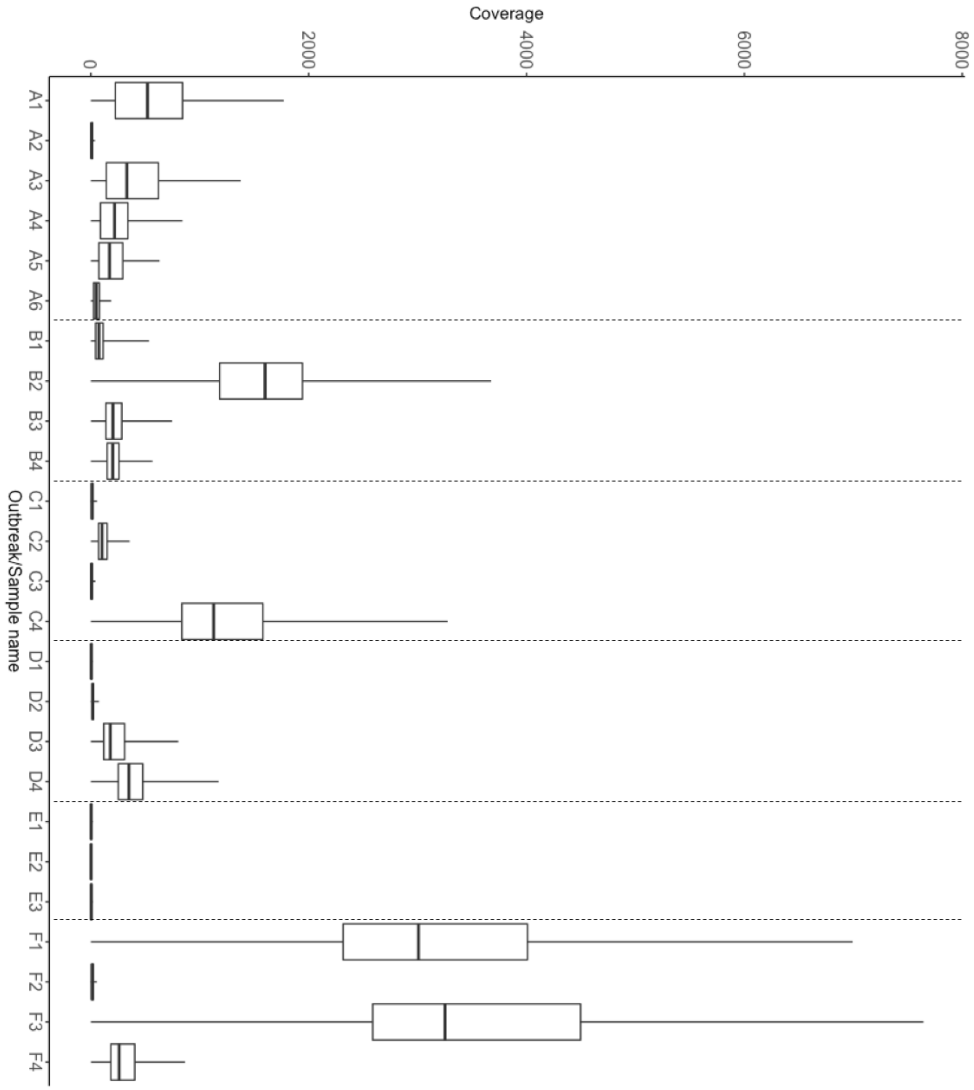


Figure 5.1 The range, interquartile range (IQR) and median of coverage for each sample analysed.

To assess whether differences in coverage between each sample were related to the HuNoV virus titre a linear regression analysis of mean coverage against the relative Ct value was performed (Figure 5.2). In this investigation, a negative trend was observed between the mean coverage and Ct value (Figure 5.2).

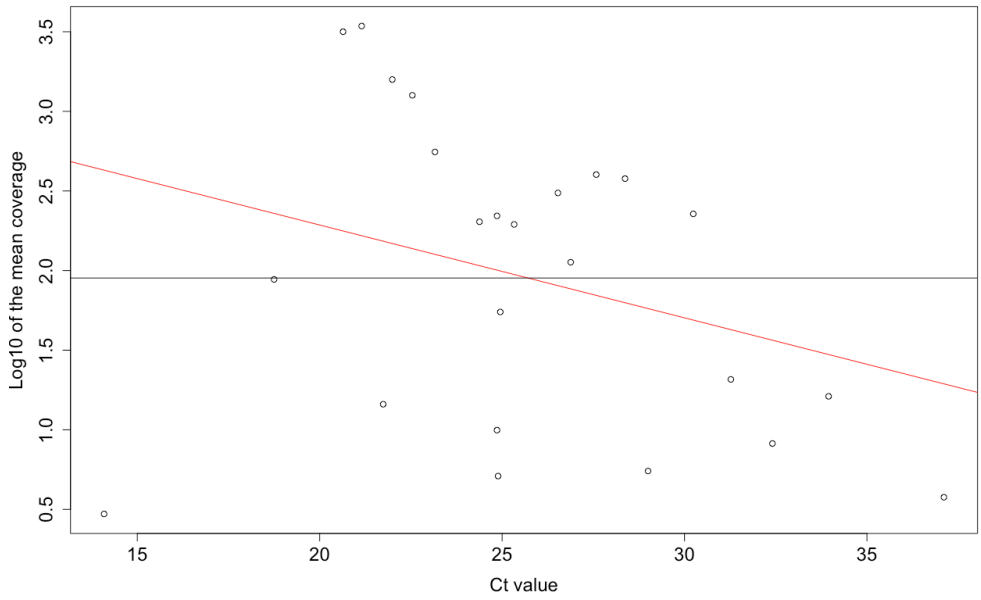


Figure 5.2 Linear regression analysis of mean sample coverage against the GII HuNoV Ct value of the RNA extract.

5.3.2 Partial Sanger sequencing of the HuNoV P2 domain was in closer agreement with its own consensus than MPS

Over half (58%) of the amplicons used in P2 domain Sanger sequencing were identical to their respective outbreak consensus sequence (Tables 5.13 - 5.18). If MPS was applied to the complete genome instead, the number of complete P2 domain matches decreased to 48% (Tables 5.13 - 5.18). Overall, the P2 domain of samples sequenced by Sanger method differed from the outbreak consensus sequence by 0-0.11%, in contrast if the P2 domain sequence was obtained via MPS this ranged from 0-3.94% (Tables 5.13 - 5.18). In the MPS data set if samples with a mean coverage below 21 were excluded the range was reduced to 0-0.58% (Tables 5.13 - 5.18). Complete genome mapping to the consensus sequence was not attained in any of the outbreaks (Tables 5.13 - 5.18).

Table 5.13 Outbreak A consensus identity measurements within partial Sanger sequencing and MPS.

Outbreak	Sample	Sanger sequence P2 identity (%)	MPS P2 identity (%)	MPS Overall identity (%)
A	1	100	100	99.34
	2	99.69	99.46	95.75
	3	100	100	99.16
	4	100	100	99.07
	5	100	100	99.02
	6	100	100	98.43

Table 5.14 Outbreak B consensus identity measurements within partial Sanger sequencing and MPS.

Outbreak	Sample	Sanger sequence P2 identity (%)	MPS P2 identity (%)	MPS Overall identity (%)
B	1	99.75	100	99.54
	2	100	100	99.90
	3	100	100	99.81
	4	100	100	99.64

Table 5.15 Outbreak C consensus identity measurements within partial Sanger sequencing and MPS.

Outbreak	Sample	Sanger sequence P2 identity (%)	MPS P2 identity (%)	MPS Overall identity (%)
C	1		99.08	96.04
	2		99.42	99.42
	3	100	97.93	93.97
	4	100	99.42	99.77

Table 5.16 Outbreak D consensus identity measurements within partial Sanger sequencing and MPS.

Outbreak	Sample	Sanger sequence P2 identity (%)	MPS P2 identity (%)	MPS Overall identity (%)
D	1	99.88	99.79	88.01
	2	99.58	98.09	96.57
	3	98.92	99.79	99.70
	4	99.78	99.79	99.77

Table 5.17 Outbreak E consensus identity measurements within partial Sanger sequencing and MPS.

Outbreak	Sample	Sanger sequence P2 identity (%)	MPS P2 identity (%)	MPS Overall identity (%)
E	1	100	99.90	99.64
	2	99.89	95.95	65.14
	3	99.90	98.44	91.03

Table 5.18 Outbreak F consensus identity measurements within partial Sanger sequencing and MPS.

Outbreak	Sample	Sanger sequence P2 identity (%)	MPS P2 identity (%)	MPS Overall identity (%)
F	1		100	99.94
	2		99.68	96.01
	3		100	99.79
	4		100	99.94

5.3.3 MPS data agreed closely with partial Sanger sequences at sufficient mean coverage

In a pairwise comparison of all outbreaks, there was > 98% agreement between the P2 domain obtained by MPS and the partial Sanger sequencing with four exceptions (Tables 5.19 – 5.23). The four exceptions were outbreak C sample 3 (97.22%), outbreak D sample 2 (97.88%), outbreak E sample 2 (88.91%) and outbreak E sample 3 (97.92%) for which the mean coverage ranged from 2.96-20.72 (Tables 5.21 and 5.23). Interestingly, there were samples with similar levels of coverage where heterogeneity between the two techniques at the P2 domain was not observed. In outbreak A and B, conserved mismatches were observed between MPS and Sanger method that were near the 5' and 3' end of the amplicon (data not shown).

Table 5.19 Outbreak A mismatch identity of the P2 domain between Sanger method and MPS. The fraction represents the number of matching base calls in the MPS data out of the total number of base calls in the Sanger sequence data.

Outbreak	Sample	Sanger sequence versus MPS identity at P2
A	1	319/321 (99.4%)
	2	315/320 (98.4%)
	3	318/320 (99.4%)
	4	319/322 (99.1%)
	5	319/322 (99.1%)
	6	317/319 (99.4%)

Table 5.20 Outbreak B mismatch identity of the P2 domain between Sanger method and MPS (Grey shading = identical). The fraction represents the number of matching base calls in the MPS data out of the total number of base calls in the Sanger sequence data.

Outbreak	Sample	Sanger sequence versus MPS identity at P2
B	1	399/399 (100%)
	2	577/578 (99.8%)
	3	578/580 (99.7%)
	4	577/578 (99.8%)

Table 5.21 Outbreak C mismatch identity of the P2 domain between Sanger method and MPS (Cross hatching = Sanger sequence data not available). The fraction represents the number of matching base calls in the MPS data out of the total number of base calls in the Sanger sequence data.

Outbreak	Sample	Sanger sequence versus MPS identity at P2
C	1	
	2	
	3	489/503 (97.2%)
	4	516/518 (99.6%)

Table 5.22 Outbreak D mismatch identity of the P2 domain between Sanger method and MPS (Grey shading = identical). The fraction represents the number of matching base calls in the MPS data out of the total number of base calls in the Sanger sequence data.

Outbreak	Sample	Sanger sequence versus MPS identity at P2
D	1	407/408 (99.8%)
	2	462/472 (97.9%)
	3	413/417 (99.0%)
	4	459/459 (100%)

Table 5.23 Outbreak E mismatch identity of the P2 domain between Sanger method and MPS (Grey shading = identical). The fraction represents the number of matching base calls in the MPS data out of the total number of base calls in the Sanger sequence data.

Outbreak	Sample	Sanger sequence versus MPS identity at P2
E	1	453/453 (100%)
	2	393/442 (88.9%)
	3	471/481 (97.9%)

5.3.4 Outbreaks clustered in an identical manner across techniques if MPS coverage was sufficient

To examine whether the sequencing technique used, or the completeness of the HuNoV genome, affected the phylogenetic relationship inferred within outbreaks the Maximum-Likelihood algorithm was applied to each genotype after alignment (Figures 5.3 – 5.7). The samples from Outbreak A, B, C and F clustered together regardless of whether the Sanger method or MPS was used to discern the complete or partial genomic sequence (Figures 5.3 – 5.7). However, if the complete genome sequence was available via MPS, phylogenetic relationships between samples in an outbreak could be inferred at greater sensitivity (Figures 5.3 – 5.7). HuNoVs from outbreaks C, D and E had sufficient similarity to cluster as a group near to the GII.4 Sydney 2012 strain in each phylogenetic tree (Figures 5.4). However, in agreement

with the mismatch analysis between Sanger method and MPS, those samples with low coverage and < 98 % similarity did not remain in the clusters assigned by partial amplicon sequencing of the P2 domain (Figures 5.4 – 5.6).

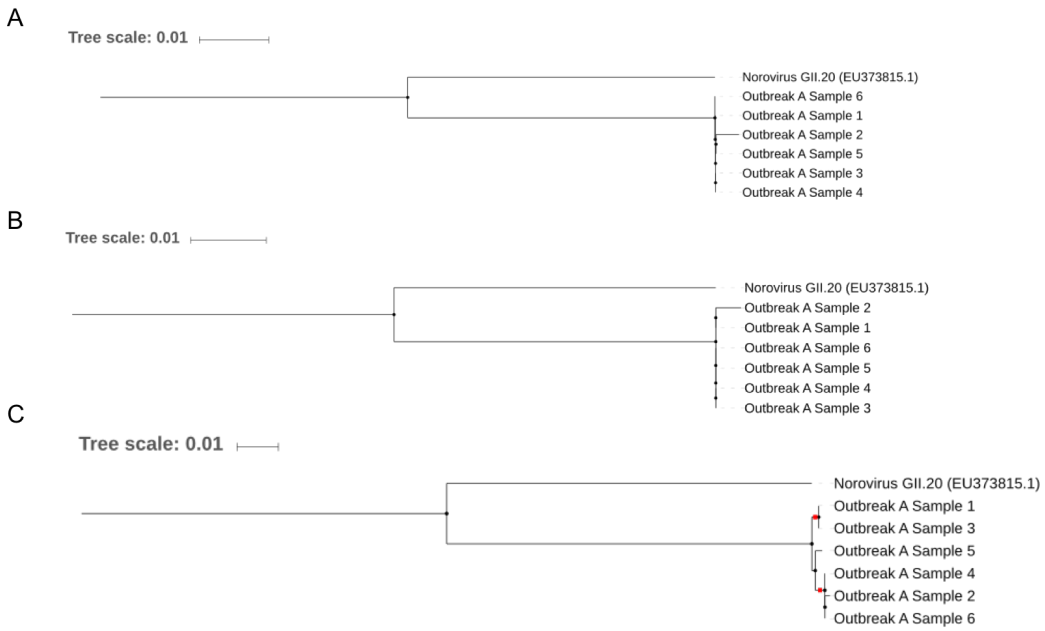


Figure 5.3 Outbreak A Maximum-Likelihood phylogenetic trees (A=Sanger method of the P2 domain, B=MPS of the P2 domain, C=MPS of the complete genome, boot strap values of nodes over 0.8 of 1000 bootstrap replicates are shown as red squares).

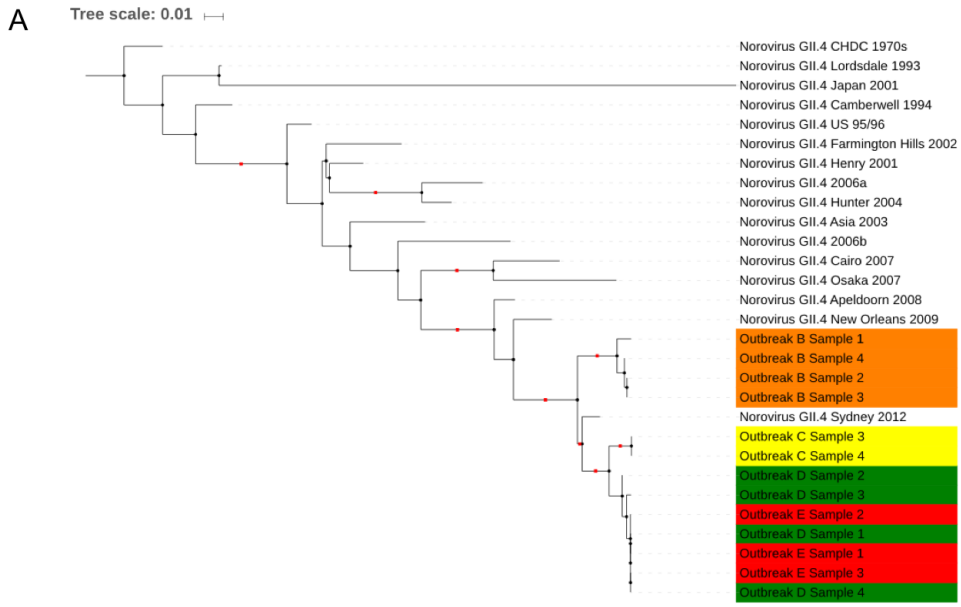


Figure 5.4 Outbreak B, C, D and E Maximum-Likelihood phylogenetic tree of P2 domain Sanger method (boot strap values of nodes over 0.8 of 1000 bootstrap replicates are shown as red squares).

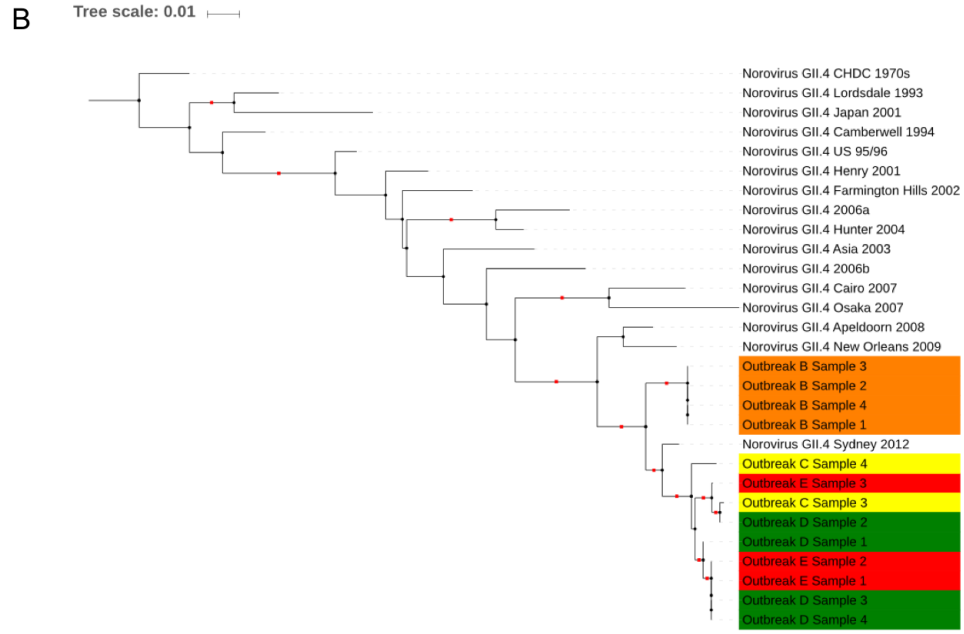


Figure 5.5 Outbreak B, C, D and E Maximum-Likelihood phylogenetic tree of P2 domain MPS (boot strap values of nodes over 0.8 of 1000 bootstrap replicates are shown as red squares).

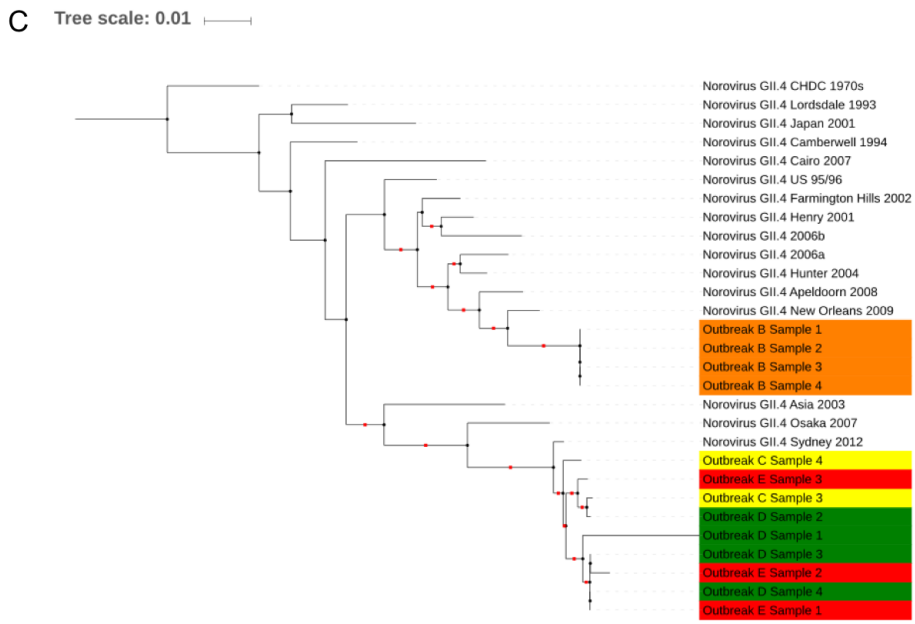


Figure 5.6 Outbreak B, C, D and E Maximum-Likelihood phylogenetic tree of complete genome MPS (boot strap values of nodes over 0.8 of 1000 bootstrap replicates are shown as red squares).

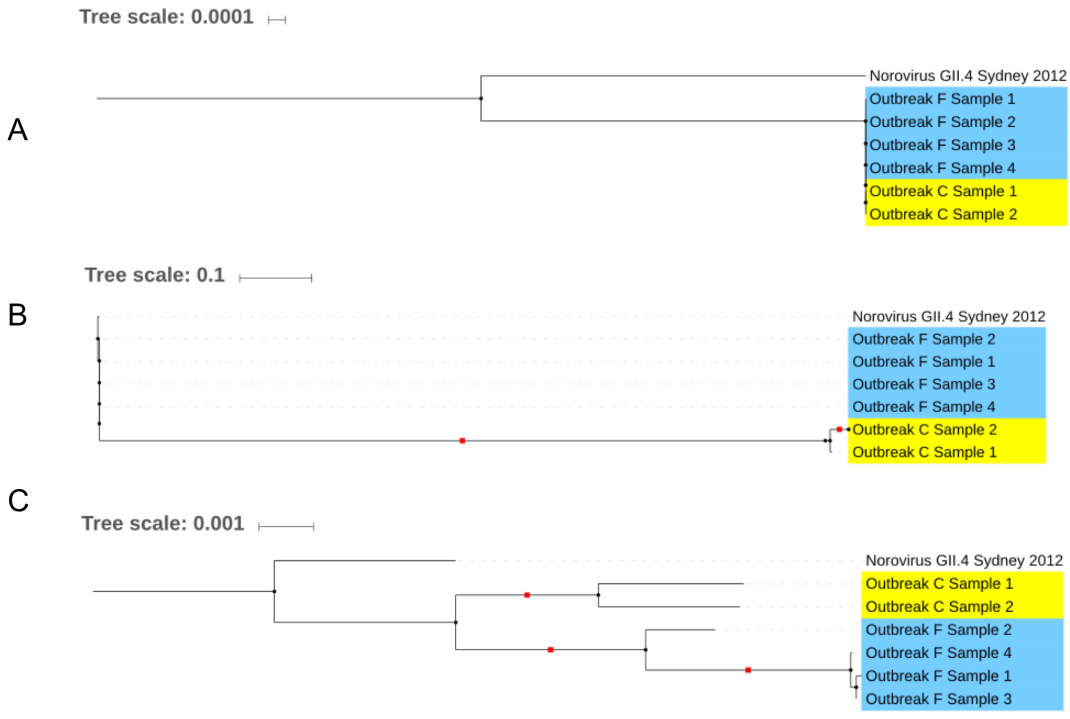


Figure 5.7 Outbreak C and F Maximum-Likelihood phylogenetic trees (A=Sanger method of region C, B=MPS of region C, C=MPS of the complete genome, boot strap values of nodes over 0.8 of 1000 bootstrap replicates are shown as red squares).

5.3.5 MPS offers a tool to detect differential selection in hosts

Minority variants were identified by applying the criteria described in Chapter 4. In total 45 minor mutations were present in the HuNoV populations of outbreaks A-F (Table 5.24). It was not possible to detect minority variants in outbreak E, due to the low coverage (Table 5.11). Interestingly, outbreak B did have sufficient coverage, but far fewer minority variants were called in contrast to the other outbreaks (Table 5.24).

Table 5.24 The frequency of minority variants identified in each ORF from outbreaks A-F.

ORF	Outbreak					
	A	B	C	D	E	F
1	10	0	7	8	0	2
2	6	0	2	7	0	2
3	0	1	0	0	0	0
Total	16	1	9	15	0	4

Most of the minority variants, across all outbreaks, did not correspond to an identical change in the consensus sequence of a separate individual (42/45) (Figures 5.8 – 5.11). However, there were three exceptions, in outbreak A, position 2011 and 5137 (Figure 5.8), and in outbreak F, position 4072 (Figure 5.11). If outbreak A was compared to the other outbreaks there were 2 positions (position 4941 and 5137) that demonstrated a similar heterogeneous population of minority variants across different individuals (Figure 5.8).

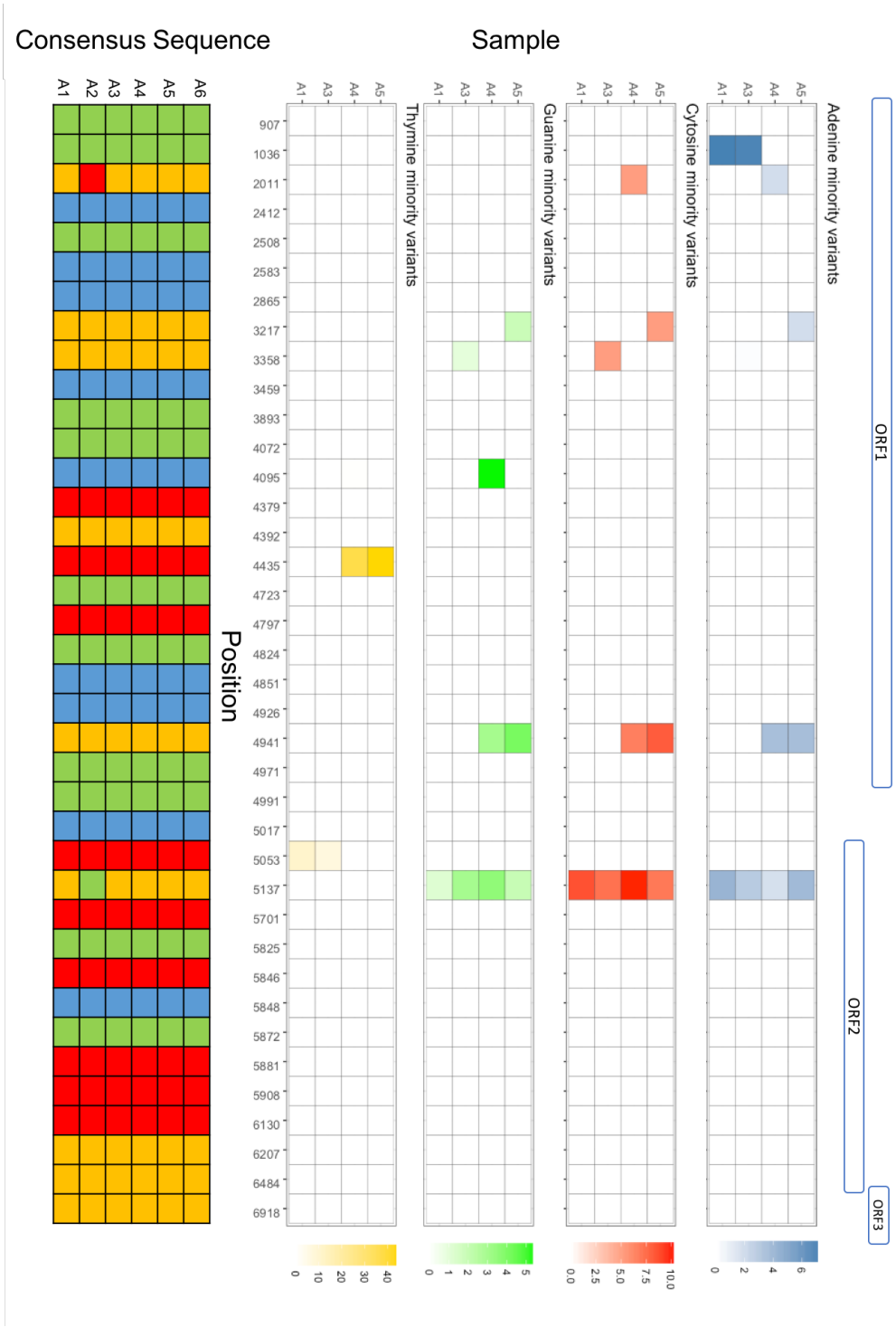


Figure 5.8 The frequency of minority variants identified in stool samples from outbreak A compared to each consensus sequence (Blue = Adenine, Red = Cytosine, Green = Guanine, Yellow = Thymine).

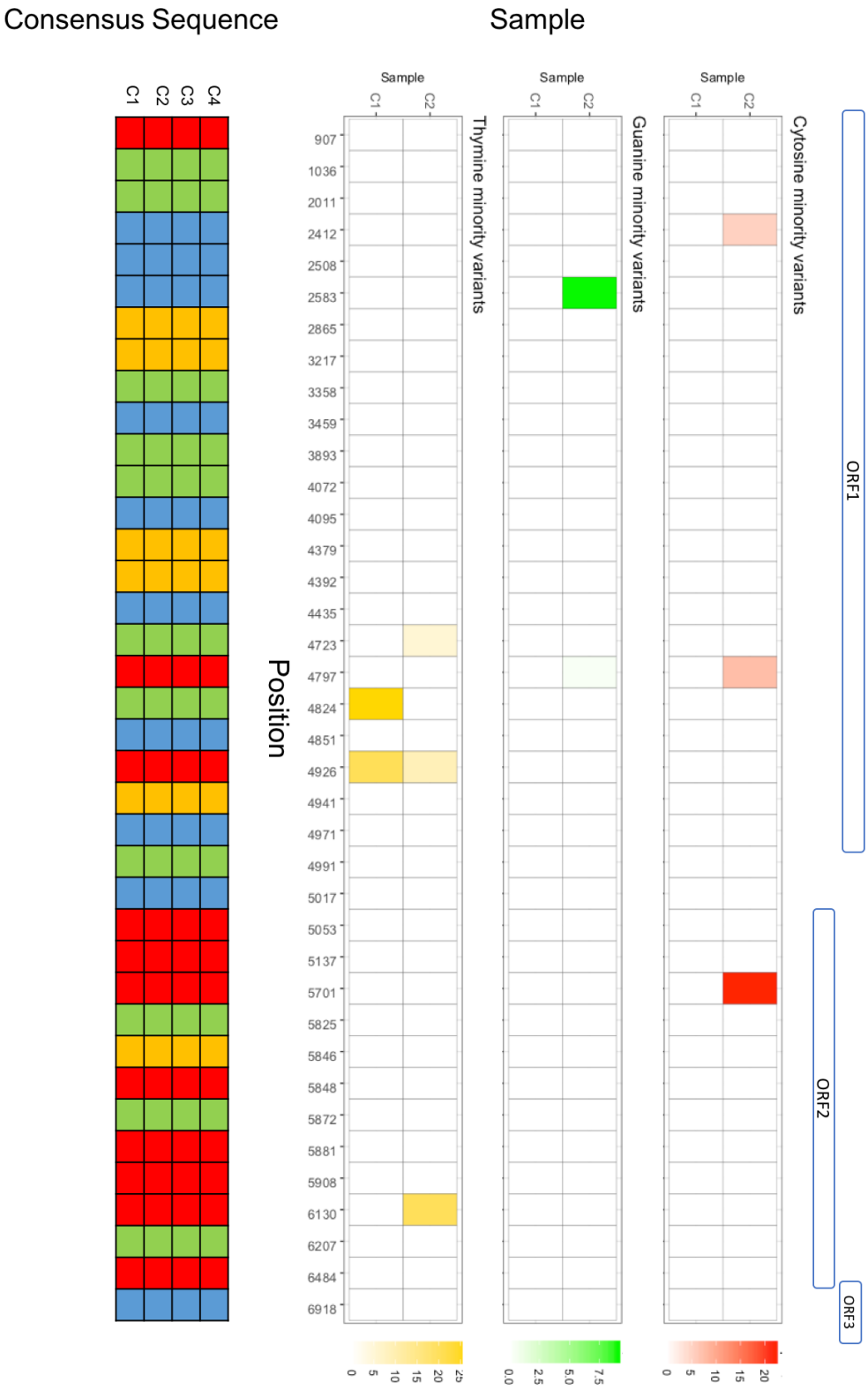


Figure 5.9 The frequency of minority variants identified in stool samples from outbreak C compared to each consensus sequence (Blue = Adenine, Red = Cytosine, Green = Guanine, Yellow = Thymine).

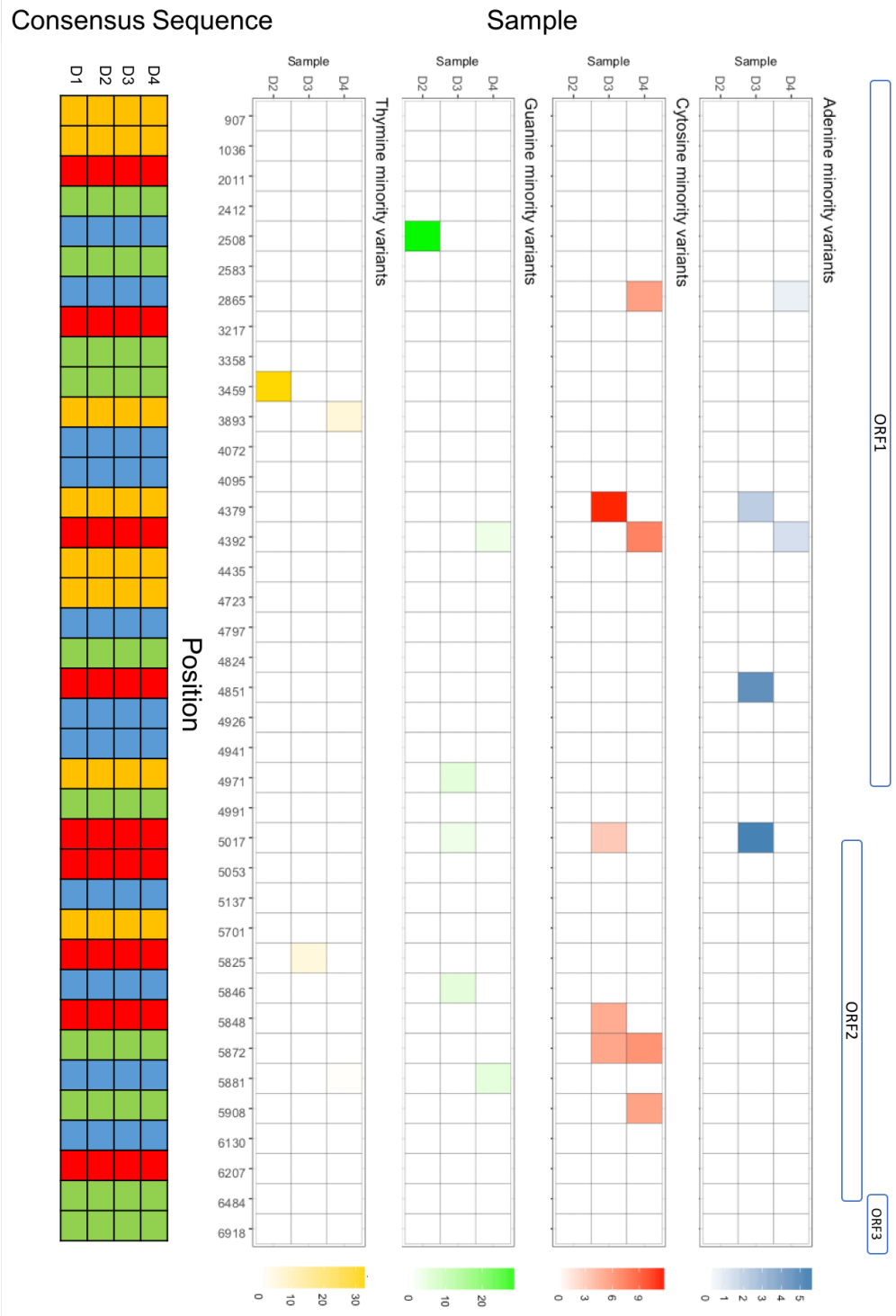


Figure 5.10 The frequency of minority variants identified in stool samples from outbreak D compared to each consensus sequence (Blue = Adenine, Red = Cytosine, Green = Guanine, Yellow = Thymine).

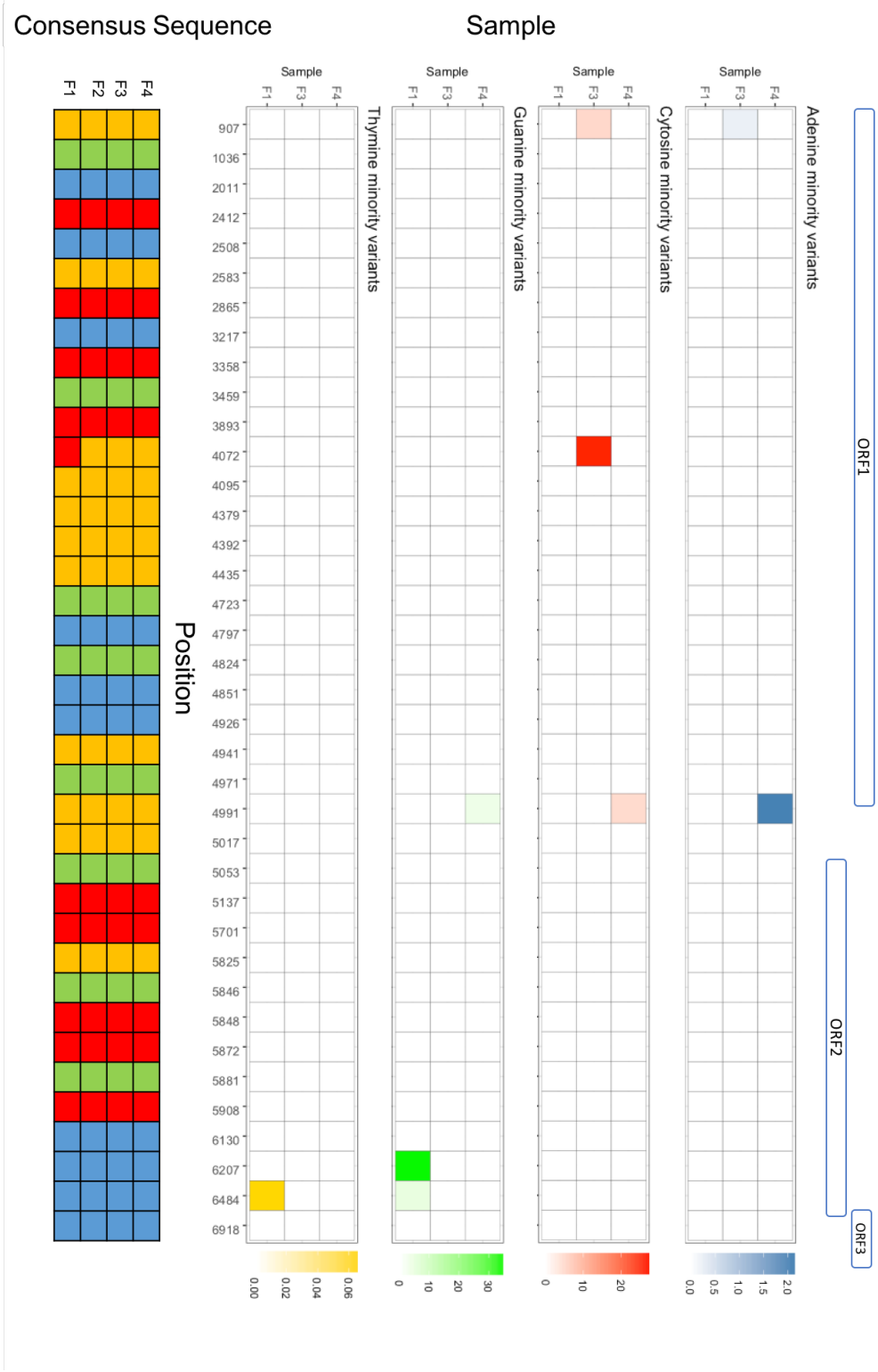


Figure 5.11 The frequency of minority variants identified in stool samples from outbreak F compared to each consensus sequence (Blue = Adenine, Red = Cytosine, Green = Guanine, Yellow = Thymine).

5.4 Discussion

5.4.1 Clinical samples with higher HuNoV viral loads tend to generate a higher mean coverage

This investigation has demonstrated that stool samples with lower HuNoV Ct values (higher viral loads) tend to generate a higher mean coverage of the virus genome, and has implications for data analysis and the ability to discriminate minority variants from error. Although this observation was not significant, it agrees with the findings of other research groups that have applied MPS to HuNoV clinical samples (Fonager et al. 2017; Nasheri et al. 2017), and it remains a major limitation. Fonager and colleagues demonstrated a strong negative correlation between the HuNoV Ct value after poly-(A) capture and the number of HuNoV reads present prior to sequencing (Fonager et al. 2017). Similarly, a strong positive correlation was shown to exist between the virus titre and number of reads mapped to the HuNoV genome without a capture method (Nasheri et al. 2017). There are three potential reasons why the correlation observed in this investigation was weak in comparison. Fonager and colleagues captured HuNoV reads after extraction, when the sample would be less complex, and pooled their nucleic acid material directly after capture which would exclude any PCR selection bias (Fonager et al. 2017). In this study, PGM-MBs were used to capture HuNoV in the clinical sample, but Nasheri and colleagues did not use a capture method which meant HuNoV saturation could not occur if the virus titre is too high (Nasheri et al. 2017).

5.4.2 Partial Sanger sequencing of the P2 domain is a valid tool to discern separate HuNoV outbreaks

A pairwise comparison between partial sequencing of the P2 domain and MPS has proved that Sanger method is an equally valid tool, if not more accurate, to distinguish separate HuNoV outbreaks.

The efficacy of partial P2 domain sequencing in HuNoV outbreaks has been established across multiple environments in the literature. If the definition, proposed by Xerry and colleagues, of a common source outbreak being 100% identical in the P2 domain is applied then 58% of samples, by Sanger sequencing, rather than 48%, by MPS were attributed to a common source (Xerry et al. 2008). Other outbreak studies have proposed this criterion is too stringent, and could misinterpret prolonged outbreaks or asymptomatic transmission (Morillo et al. 2017; Sukhrie et al. 2013). If Sanger or MPS was applied, samples from each outbreak clustered together spatio-temporally, but limited depth in MPS data or mismatches outside of the P2 domain led to phylogenetic rearrangements within clusters. It is also possible the phylogenetic rearrangements observed during the analysis of sequence data outside of the P2 domain (i.e Outbreak C – Sample 3 and 4) could have been caused by recombination, the presence of two different HuNoVs or cross-contamination during the library preparation stage (Figures 5.4 – 5.6).

Barrabeig and colleagues previously targeted a highly-conserved site of a GII.2 HuNoV genome (region C) to implicate an asymptomatic food handler in a point source outbreak (Barrabeig et al. 2010). However, genomic regions that are more variable will provide greater sensitivity in detecting genetic linkage of transmission

between individual hosts. The difference in sensitivity between the P2 subdomain of HuNoV and region C was proved when partial sequences of each site from stool samples in two immunocompromised individuals were compared to environmental swabs (Xerry et al. 2010). Although, discernment of nosocomial transmission from novel HuNoV virus introductions has been achieved by targeting region C in a hospital outbreak previously (Rahamat-Langendoen et al. 2013).

5.4.3 MPS can provide evidence of differential selection in the host and increase the resolution of transmission within an outbreak

This study has however highlighted the advantages of deep sequencing, over Sanger sequencing, by providing complete resolution of the HuNoV genome in each stool sample, and detecting minority variants which can further characterize differences in virus populations of individual hosts. The advantage of MPS in HuNoV outbreak control has been described previously, where the 454 (Roche) instrument was used to show nosocomial transmission of minority alleles from one immunosuppressed patient to another, which then became predominant in the latter (Kundu et al. 2013). In a similar manner, outbreak F occurred in a hospital, and a minor allele (At position 4072) observed in an individual was observed to be the dominant allele in a separate individual. However, without further epidemiological data it can only be speculated whether this was due to a bottleneck transmission event or exposure to a common source of infection and differential selection in each host. The minor allele present in individual F3, provided further support to the MPS P2 subdomain sequence data that F1, F3 and F4 were from the same HuNoV introduction, even though Sanger sequencing had failed. It should be

noted that it is possible to detect minority variants in chromatograms generated by Sanger sequencing technology, however, the minor allele must be present above a greater threshold in contrast to deep sequencing (Larder et al. 1993; Palmer et al. 2005).

Outbreak A occurred in a restaurant, and in this suspected outbreak by applying the 100% criteria on P2 subdomain identity to the MPS or Sanger sequence consensus data, this would have been considered as more than one virus introduction, with five identical strains constituting one outbreak and another discrete event with a single case. However, the ability to obtain greater coverage and the identification of minority variants highlighted the presence of one variant (position 5137) in more than one sample and another variant present in one other sample (position 2011), supporting the notion that all six strains were linked and constituted the same outbreak. The observed change, in individual A2, for which minority variants could not be detected due to insufficient coverage could be further evidence of differential selection.

A retrospective MPS investigation of a presumed GII.Pg/GII.3 recombinant point source outbreak in 1972 was recently performed with the Ion torrent instrument (Johnson et al. 2017). Johnson and colleagues described the presence of significant subpopulation heterogeneity in the GIIPg/GII.3 recombinant population across several individuals at amino acid positions 315 and 1293, which was proposed to be a consequence of the same virus population infecting separate hosts (Johnson et al. 2017). In this investigation, a heterogeneous subpopulation was present in outbreak A across 4/6 individuals at position 5137, which could be due to exposure

to a common infection source. To a lesser extent there was a heterogeneous subpopulation at position 4941 in 2/6 individuals. The capability of HuNoV P2 domain Sanger sequencing to track virus transmission between individuals was evident when a predominant mutation, that appeared in a chronically infected individual, was detected 6-month-old child (Sukhrie et al. 2010). However, P2 domain sequencing would be less sensitive to mutations that occur outside of the hypervariable region, which could be more likely if the virus is not under selective pressure.

MPS and partial Sanger sequencing of the P2 domain offer two separate methodologies to discern separate HuNoV introduction events and different outbreaks. If sufficient depth of MPS sequence data was available, then agreement existed between both techniques phylogenetically. However, from a public health perspective, Sanger sequencing would have a higher rate of agreement toward the respective outbreak consensus sequence (58%) at a shorter turnaround time and reduced cost, which is often necessary to implement interventions and control the spread of illness. If MPS is applied, a narrow window of time will exist at which individuals have not reached the convalescent stages of illness and the viral load in clinical samples is high. Those stool samples with the highest viral loads can provide the complete HuNoV sequence at high accuracy, and minority variants can be reliably called, to further strengthen confidence in individual transmission events if they correspond to minority or dominant alleles in other infected individuals. This advantage has already been demonstrated in HCV infection, as Montoya and colleagues shown linkages could be made between minority variants of NS5 gene

that were not detectable by Sanger sequencing (Montoya et al. 2016). In this study, four minority variants were detected in four individual hosts that supported the formation of two additional clusters in comparison to Sanger sequencing (Montoya et al. 2016).

6 Final discussion and conclusions

A highly reproducible quantitative assay for human norovirus (HuNoV) in stool samples has been optimized. The approach developed and optimized in this thesis has several clinical applications: in monitoring progress of disease, establishment of standardized cut off values from discrimination between symptomatic and asymptomatic infection, based on viral RNA copies rather than assay Ct values which is subject to significant lab to lab variability. This work confirmed the unreliability of the use of a house keeping gene in sample standardization and relative quantification, but demonstrated that the expression of copy numbers in relation to stool weight yielded highly reproducible results in repeated experiments, dispelling concerns around the lack of homogenous distribution of the virus in a stool sample. Furthermore, the development of an RNA standard demonstrated that currently widely used cDNA standards (ISO Method for norovirus detection in food and the environment) significantly and systematically underestimated the amount of virus present in a sample. This is an important consideration, when viral load may be used as a proxy for determining the risk a contaminated food item may pose to consumers, and suggests recalibration of such standards and assumptions on viral load and contamination may be necessary.

Further work should examine how effective this methodology is in the quantification of other genotypes, for instance whether under this methodology the GII specific qPCR is equally efficient and sensitive in amplifying other genotypes from the same group. Similarly, it may also be interesting and useful to validate

these findings by designing and producing a genogroup I specific standard that can be used to quantify GI HuNoVs, which are frequently associated with food borne infections.

This optimized HuNoV quantification protocol was subsequently applied for the optimization of the capture/enrichment and MPS sequencing methods developed and described in this thesis. Furthermore, the protocol is also being successfully applied for monitoring viral load in a chronically infected patient undergoing experimental treatment with Ribavirin (this ongoing work is not part of this thesis).

Porcine gastric mucin-Magnetic beads (PGM-MBs) were used for the first time to capture HuNoV from clinical samples prior to Massively Parallel Sequencing (MPS) as a method for HuNoV sequence enrichment. Although the PGM-MB capture methods showed inter-stool suspension variability that did not always correlate with viral load, it allowed the recovery of near complete genomes in most samples even from those collected post-acute phase of disease, where viral loads were low. A tendency for mean coverage to decrease in the samples collected in the convalescent phase was found, coupled with an inability to use the capture method to efficiently serve as a virus concentration method. It is possible that in these samples, virus capture with PGM may be limited by the presence of copro-antibodies bound to the virus, as subsequently seen by performing HuNoV VLP-specific EIA. Therefore, to further improve the capture method to allow virus concentration in low viral load samples, further investigations should examine the potential for use of detergents, acid solutions or light sonication to dissociate virus-

antibody complexes prior to PGM-MB capture, which would increase accessibility to aggregated or antibody (Ab) neutralized virus.

The developed capture method provided an opportunity to apply MPS to HuNoV virus populations of acute infections in immunocompetent individuals across more time points, and over a longer period than any previous studies have described. A pipeline was developed to call minority variants and two separate methods were applied to discriminate true positives from sequencing error, which led to further proof of the existence of quasispecies in acute infection. This is the first study to link the presence of coproantibodies to the presence of minority variants or dominant nonsynonymous mutations in the HuNoV major capsid protein, and further, map these changes to the virus surface. Comparisons made between a small subset of epochal VLPs demonstrated a potential role for Original Antigenic Sin (OAS) or Antigenic Seniority in coproantibody responses to HuNoV, and the presence of a virus, closely related to the Sydney strain, for which the Ab response present in the acute phase appeared to be ineffective. However, it was noted that an arbitrary cut-off for minority variants is accurate but insensitive, and instead further experiments should determine true minority variants through repeat experiments or an RNA sequencing internal control. Moreover, if a minority variant exists it should exist at a similar frequency among the replicates, and therefore further investigations should assess whether this is true. Those mapped amino acids (AAs) considered to be of antigenic importance in this study should be further characterized by site directed mutagenesis of VLPs to further strengthen these observations.

This study is the first direct comparison between MPS and Sanger sequencing, and how each technique can influence how a HuNoV outbreak/transmission event is interpreted, with potential impact on infection control and public health actions. The presence of minority variants strengthened phylogenetic relationships in common source outbreaks, when the consensus sequence would suggest otherwise. But given the variability in MPS coverage that can occur and the slow turn over time in contrast to Sanger sequencing, the latter can still be considered a more suitable or pragmatic public health tool. Further investigations should examine whether 3rd generation sequencing technologies (i.e MinION) are more reasonable tools for HuNoV outbreaks, as these also have added benefits such as lower cost and near-real time processing.

7 References

Abbate I, Rozera G, Tommasi C, Bruselles A, Bartolini B, Chillemi G, et al. Analysis of co-receptor usage of circulating viral and proviral HIV genome quasispecies by ultra-deep pyrosequencing in patients who are candidates for CCR5 antagonist treatment. Clin. Microbiol. Infect. 2011;17(5):725–31.

Adler, J. L. and RZ. Winter Vomiting Disease. J. Infect. Dis. 1969;119:668–673.

Agus SG, Dolin R, Wyatt RG. Acute infectious nonbacterial gastroenteritis: intestinal histopathology. Histologic and enzymatic alterations during illness produced by the Norwalk agent in man. Ann. Intern. Med. 1973. p. 18–25.

Ahmed SM, Hall AJ, Robinson AE, Verhoef L, Premkumar P, Parashar UD, et al. Global prevalence of norovirus in cases of gastroenteritis: a systematic review and meta-analysis. Lancet Infect Dis [Internet]. 2014;14(8):725–30. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24981041

Aird D, Ross MG, Chen W-S, Danielsson M, Fennell T, Russ C, et al. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol. [Internet]. 2011;12(2):R18. Available from: http://genomebiology.biomedcentral.com/articles/10.1186/gb-2011-12-2-r18

Allen DJ, Gray JJ, Gallimore CI, Xerry J, Iturriza-Gomara M. Analysis of amino acid variation in the P2 domain of the GII-4 norovirus VP1 protein reveals putative variant-specific epitopes. PLoS One [Internet]. 2008;3(1):e1485. Available from: https://www.ncbi.nlm.nih.gov/pubmed/18213393

Allen DJ, Noad R, Samuel D, Gray JJ, Roy P, Iturriza-Gómara M. Characterisation of a GII-4 norovirus variant-specific surface-exposed site involved in antibody binding. Virol. J. [Internet]. 2009;6(1):150. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2762976&tool=pmcentrez&rendertype=abstract

Amar CF, East CL, Gray J, Iturriza-Gomara M, Maclure EA, McLauchlin J. Detection by PCR of eight groups of enteric pathogens in 4,627 faecal samples: re-examination of the English case-control Infectious Intestinal Disease Study (1993-1996). Eur J Clin Microbiol Infect Dis [Internet]. 2007;26(5):311–23. Available from: https://www.ncbi.nlm.nih.gov/pubmed/17447091

Andino R, Domingo E. Viral quasispecies. Virology. 2015. p. 46–51.

Ando T, Monroe SS, Gentsch JR, Jin Q, Lewis DC, Glass RI. Detection and differentiation of antigenically distinct small round- structured viruses (Norwalk-like viruses) by reverse transcription-PCR and southern hybridization. J. Clin. Microbiol. 1995;33(1):64–71.

Ando T, Noel JS, Fankhauser RL. Genetic Classification of " Norwalk-like Viruses ." 2000;30333(Suppl 2):336–49.

Andrews S. FastQC: A quality control tool for high throughput sequence data. Http://Www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/. 2010. p. http://www.bioinformatics.babraham.ac.uk/projects/.

Angly FE, Willner D, Prieto-Davó A, Edwards RA, Schmieder R, Vega-Thurber R, et al. The GAAS metagenomic tool and its estimations of viral and microbial average genome size in four major biomes. PLoS Comput. Biol. 2009;5(12).

Anoop M, Mathew AJ, Jayakumar B, Issac A, Nair S, Abraham R, et al. Complete genome sequencing and evolutionary analysis of dengue virus serotype 1 isolates from an outbreak in Kerala, South India. Virus Genes. 2012;45(1):1–13.

Archer J, Baillie G, Watson SJ, Kellam P, Rambaut A, Robertson DL. Analysis of high-depth sequence data for studying viral diversity: A comparison of next generation sequencing platforms using Segminator II. BMC Bioinformatics. 2012;13(1).

Arezi B, Hogrefe HH. Escherichia coli DNA polymerase III epsilon subunit increases Moloney murine leukemia virus reverse transcriptase fidelity and accuracy of RT-PCR procedures. Anal Biochem [Internet]. 2007;360(1):84–91. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17107651

Arias A, Thorne L, Ghurburrun E, Bailey D, Goodfellow I. Norovirus Polymerase Fidelity Contributes to Viral Transmission In Vivo. mSphere [Internet]. 2016;1(5). Available from: https://www.ncbi.nlm.nih.gov/pubmed/27777985

Arias A, Thorne L, Goodfellow I. Favipiravir elicits antiviral mutagenesis during virus replication in vivo. Elife [Internet]. 2014;3:e03679. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25333492

Atmar RL, Bernstein DI, Harro CD, Al-Ibrahim MS, Chen WH, Ferreira J, et al. Norovirus vaccine against experimental human Norwalk Virus illness. N Engl J Med [Internet]. 2011;365(23):2178–87. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22150036

Atmar RL, Bernstein DI, Lyon GM, Treanor JJ, Al-Ibrahim MS, Graham DY, et al. Serological Correlates of Protection against a GII.4 Norovirus. Clin Vaccine Immunol [Internet]. 2015;22(8):923–9. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26041041

Atmar RL, Opekun AR, Gilger M a, Estes MK, Crawford SE, Neill FH, et al. Norwalk virus shedding after experimental human infection. Emerg. Infect. Dis. [Internet]. 2008 Oct [cited 2014 Oct 23];14(10):1553–7. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2609865&tool=pmcentrez&rendertype=abstract

Atmar RL, Opekun AR, Gilger M a, Estes MK, Crawford SE, Neill FH, et al. Determination of the 50% human infectious dose for Norwalk virus. J. Infect. Dis. [Internet]. 2014 Apr 1 [cited 2014 Nov 6];209(7):1016–22. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24253285

Baize S, Pannetier D, Oestereich L, Rieger T, Koivogui L, Magassouba N, et al. Emergence of Zaire Ebola Virus Disease in Guinea. N. Engl. J. Med. [Internet]. 2014;371(15):1418–25. Available from: http://www.nejm.org/doi/10.1056/NEJMoa1404505

Baltimore D. Expression of Animal Virus Genomes. Am. Soc. Microbiol. 1971;35(3):235–41.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J. Comput. Biol. [Internet]. 2012;19(5):455–77. Available from: http://online.liebertpub.com/doi/abs/10.1089/cmb.2012.0021

Bar-Eli M, Ahuja H, Foti A, Cline MJ. N-RAS mutations in T-cell acute lymphocytic leukaemia: analysis by direct sequencing detects a novel mutation. Br J Haematol [Internet]. 1989;72(1):36–9. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=2660900

Barclay L, Park GW, Vega E, Hall A, Parashar U, Vinje J, et al. Infection control for norovirus. Clin Microbiol Infect [Internet]. 2014;20(8):731–40. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24813073

Baric R, Yount B, Lindesmith L. Expression and self-assembly of Norwalk virus capsid protein from Venezuelan equine encephalitis virus replicons. J. Virol. [Internet]. 2002;76(6):3023–30. Available from: http://jvi.asm.org/content/76/6/3023.short

Barrabeig I, Rovira A, Buesa J, Bartolome R, Pinto R, Prellezo H, et al. Foodborne norovirus outbreak: the role of an asymptomatic food handler. BMC Infect Dis [Internet]. 2010;10:269. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20843351

Bartolini B, Giombini E, Zaccaro P, Selleri M, Rozera G, Abbate I, et al. Extent of HCV NS3 protease variability and resistance-associated mutations assessed by next generation sequencing in HCV monoinfected and HIV/HCV coinfected patients. Virus Res. 2013;177(2):205–8.

Bartsch SM, Lopman BA, Hall AJ, Parashar UD, Lee BY. The potential economic value of a human norovirus vaccine for the United States. Vaccine [Internet]. 2012;30(49):7097–104. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23026689

Bartsch SM, Lopman BA, Ozawa S, Hall AJ, Lee BY. Global Economic Burden of Norovirus Gastroenteritis. PLoS One [Internet]. 2016;11(4):e0151219. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27115736

Batty EM, Wong TH, Trebes A, Argoud K, Attar M, Buck D, et al. A modified RNA-Seq approach for whole genome sequencing of RNA viruses from faecal and blood samples. PLoS One [Internet]. 2013;8(6):e66129. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23762474

Bavelaar HH, Rahamat-Langendoen J, Niesters HG, Zoll J, Melchers WJ. Whole genome sequencing of fecal samples as a tool for the diagnosis and genetic characterization of norovirus. J Clin Virol [Internet]. 2015;72:122–5. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26492615

Belliot G, Sosnovtsev S V, Chang KO, McPhie P, Green KY. Nucleotidylylation of the VPg protein of a human norovirus by its proteinase-polymerase precursor protein. Virology [Internet]. 2008;374(1):33–49. Available from: https://www.ncbi.nlm.nih.gov/pubmed/18234264

Belliot G, Sosnovtsev S V, Mitra T, Hammer C, Garfield M, Green KY. In vitro proteolytic processing of the MD145 norovirus ORF1 nonstructural polyprotein yields stable precursors and products similar to those detected in calicivirus-infected cells. J Virol [Internet]. 2003;77(20):10957–74. Available from: https://www.ncbi.nlm.nih.gov/pubmed/14512545

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature [Internet]. 2008;456(7218):53–9. Available from:

http://www.nature.com/doifinder/10.1038/nature07517

Bermúdez de León M, Peñuelas-Urquides K, Aguado-Barrera ME, Currás-Tuala MJ, Escobedo-Guajardo BL, González-Ríos RN, et al. In vitro transcribed RNA molecules for the diagnosis of pandemic 2009 influenza A(H1N1) virus by real-time RT-PCR. J. Virol. Methods. 2013;193(2):487–91.

Bernstein DI, Atmar RL, Lyon GM, Treanor JJ, Chen WH, Jiang X, et al. Norovirus vaccine against experimental human GII.4 virus illness: a challenge study in healthy adults. J Infect Dis [Internet]. 2015;211(6):870–8. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25210140

Bertolotti-Ciarlet A, Crawford SE, Hutson AM, Estes MK. The 3' end of Norwalk virus mRNA contains determinants that regulate the expression and stability of the viral capsid protein VP1: a novel function for the VP2 protein. J. Virol. 2003;77(21):11603–15.

Bertolotti-ciarlet A, White LJ, Chen R, Prasad V, Estes MK, Prasad BVV. Structural Requirements for the Assembly of Norwalk Virus-Like Particles Structural Requirements for the Assembly of Norwalk Virus-Like Particles. J. Virol. 2002;76(8):4044–55.

Blacklow NR, Cukor G, Bedigian MK, Echeverria P, Greenberg HB, Schreiber DS, et al. Immune response and prevalence of antibody to Norwalk enteritis virus as determined by radioimmunoassay. J. Clin. Microbiol. 1979;10(6):903–9.

Blaney DD, Daly ER, Kirkland KB, Tongren JE, Kelso PT, Talbot EA. Use of alcohol-based hand sanitizers as a risk factor for norovirus outbreaks in long-term care facilities in northern New England: December 2006 to March 2007. Am J Infect Control [Internet]. 2011;39(4):296–301. Available from: https://www.ncbi.nlm.nih.gov/pubmed/21411187

Blazevic V, Malm M, Honkanen H, Knip M, Hyoty H, Vesikari T. Development and maturation of norovirus antibodies in childhood. Microbes Infect [Internet]. 2016;18(4):263–9. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26724451

Bok K, Green KY. Norovirus gastroenteritis in immunocompromised patients. N Engl J Med [Internet]. 2012;367(22):2126–32. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23190223

Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.

Boom R, Sol CJ, Salimans MMM, Jansen CL, Wertheim-van Dillen PM, van der Noordaa J, et al. Rapid and simple method for purification of nucleic acids. J. Clin. Microbiol. [Internet]. 1990;28(3):495–503. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=269651&tool=pmcentrez&rendertype=abstract

Boon D, Mahar JE, Abente EJ, Kirkwood CD, Purcell RH, Kapikian AZ, et al. Comparative evolution of GII.3 and GII.4 norovirus over a 31-year period. J Virol [Internet]. 2011;85(17):8656–66. Available from: https://www.ncbi.nlm.nih.gov/pubmed/21715504

Bowers RM, Dhar AK. Effect of template on generating a standard curve for absolute quanti fi cation of an RNA virus by real-time reverse transcriptase-polymerase chain reaction. Mol. Cell. Probes [Internet]. Elsevier Ltd; 2011;25(1):60–4. Available from: http://dx.doi.org/10.1016/j.mcp.2010.12.002

Braslavsky I, Hebert B, Kartalov E, Quake SR. Sequence information can be obtained from single DNA molecules. [Internet]. Proc. Natl. Acad. Sci. U. S. A. 2003. p. 3960–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12651960

Brouwer M, Lievens B, Hemelrijck W Van, Ackerveken G Van Den, Cammue BPA, Thomma BPHJ. Quantification of disease progression of several microbial pathogens on Arabidopsis thaliana using real-time fluorescence PCR. 2003;228:241–8.

Brown JR, Roy S, Ruis C, Yara Romero E, Shah D, Williams R, et al. Norovirus Whole-Genome Sequencing by SureSelect Target Enrichment: a Robust and Sensitive Method. J Clin Microbiol [Internet]. 2016;54(10):2530–7. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27487952

Bruggink LD, Oluwatoyin O, Sameer R, Witlox KJ, Marshall JA. Molecular and epidemiological features of gastroenteritis outbreaks involving genogroup I norovirus in Victoria, Australia, 2002-2010. J Med Virol [Internet]. 2012;84(9):1437–48. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22825823

de Bruin E, Duizer E, Vennema H, Koopmans MP. Diagnosis of Norovirus outbreaks by commercial ELISA or RT-PCR. J Virol Methods [Internet]. 2006;137(2):259–64. Available from: https://www.ncbi.nlm.nih.gov/pubmed/16901556

Bull R a., Eden J-S, Luciani F, McElroy K, Rawlinson WD, White P a. Contribution of Intra- and Interhost Dynamics to Norovirus Evolution. J. Virol. 2012;86(6):3219–29.

Bull R a., Tanaka MM, White P a. Norovirus recombination. J. Gen. Virol. 2007;88(12):3347–59.

Bull RA, Eden JS, Rawlinson WD, White PA. Rapid evolution of pandemic noroviruses of the GII.4 lineage. PLoS Pathog [Internet]. 2010;6(3):e1000831. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20360972

Bull RA, Hansman GS, Clancy LE, Tanaka MM, Rawlinson WD, White PA. Norovirus recombination in ORF1/ORF2 overlap. Emerg Infect Dis [Internet]. 2005;11(7):1079–85. Available from: https://www.ncbi.nlm.nih.gov/pubmed/16022784

Bull RA, White PA. Mechanisms of GII.4 norovirus evolution. Trends Microbiol [Internet]. 2011;19(5):233–40. Available from: https://www.ncbi.nlm.nih.gov/pubmed/21310617

Bullen TF, Forrest S, Campbell F, Dodson AR, Hershman MJ, Pritchard DM, et al. Characterization of epithelial cell shedding from human small intestine. Lab. Investig. [Internet]. 2006;86(10):1052–63. Available from: http://www.nature.com/doifinder/10.1038/labinvest.3700464

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE Guidelines : M inimum I nformation for Publication of Q uantitative Real-Time PCR E xperiments SUMMARY : 2009;622:611–22.

Cannon JL, Barclay L, Collins NR, Wikswo ME, Castro CJ, Magaña LC, et al. Genetic and epidemiologic trends of norovirus outbreaks in the United States from 2013 to 2016 demonstrated emergence of novel GII.4 eecombinant viruses. J. Clin. Microbiol. 2017;55(7):2208–21.

Cao S, Lou Z, Tan M, Chen Y, Liu Y, Zhang Z, et al. Structural basis for the recognition of blood group trisaccharides by norovirus. J. Virol. 2007;81(11):5949–57.

Carlsson B, Lindberg AM, Rodriguez-Díaz J, Hedlund KO, Persson B, Svensson L. Quasispecies dynamics and molecular evolution of human norovirus capsid P region during chronic infection. J. Gen. Virol. 2009;90(2):432–41.

Carmona-Vicente N, Vila-Vicent S, Allen D, Gozalbo-Rovira R, Iturriza-Gomara M, Buesa J, et al. Characterization of a Novel Conformational GII.4 Norovirus Epitope: Implications for Norovirus-Host Interactions. J Virol [Internet]. 2016;90(17):7703–14. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27307569

Carstens EB, Ball L a. Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2008). Arch. Virol. 2009;154(7):1181–8.

Chaipan C, Pryszlak A, Dean H, Poignard P, Benes V, Griffiths AD, et al. Single-Virus Droplet Microfluidics for High-Throughput Screening of Neutralizing Epitopes on HIV Particles. Cell Chem. Biol. 2017;24(6):751–757.e3.

Chakravarty S, Hutson AM, Estes MK, Prasad B V. Evolutionary trace residues in noroviruses: importance in receptor binding, antigenicity, virion assembly, and strain diversity. J Virol [Internet]. 2005;79(1):554–68. Available from: https://www.ncbi.nlm.nih.gov/pubmed/15596848

Chan MC, Sung JJ, Lam RK, Chan PK, Lee NL, Lai RW, et al. Fecal viral load and norovirus-associated gastroenteritis. Emerg Infect Dis [Internet]. 2006;12(8):1278–80. Available from: https://www.ncbi.nlm.nih.gov/pubmed/16965715

Chang KO, Sosnovtsev S V, Belliot G, King AD, Green KY. Stable expression of a Norwalk virus RNA replicon in a human hepatoma cell line. Virology [Internet]. 2006;353(2):463–73. Available from: https://www.ncbi.nlm.nih.gov/pubmed/16843517

Cheetham S, Souza M, Meulia T, Grimes S, Han MG, Saif LJ. Pathogenesis of a genogroup II human norovirus in gnotobiotic pigs. J Virol [Internet]. 2006;80(21):10372–81. Available from: https://www.ncbi.nlm.nih.gov/pubmed/17041218

Chen R, Neill JD, Estes MK, Prasad BVV. X-ray structure of a native calicivirus: structural insights into antigenic diversity and host specificity. Proc. Natl. Acad. Sci. U. S. A. 2006;103(21):8048–53.

Chen R, Neill JD, Noel JS, Anne M, Glass RI, Estes MK, et al. Inter- and Intragenus Structural Variations in Caliciviruses and Their Functional Implications Inter- and Intragenus Structural Variations in Caliciviruses and Their Functional Implications. J. Virol. 2004;78(12):6469–79.

Chi CY, Tseng FC, Liu DP, Chang YW, Wu HC, Huang YF, et al. Investigations of clinical isolations of oral poliovirus vaccine strains between 2000 and 2005 in southern Taiwan. J. Clin. Virol. 2009;45(2):129–34.

Choi YS, Koo ES, Kim MS, Choi JD, Shin Y, Jeong YS. Re-emergence of a GII.4 Norovirus Sydney 2012 Variant Equipped with GII.P16 RdRp and Its Predominance over Novel Variants of GII.17 in South Korea in 2016. Food Env. Virol [Internet]. 2017; Available from: https://www.ncbi.nlm.nih.gov/pubmed/28120262

Chong Y, Ikematsu H. Effect of seasonal vaccination on the selection of influenza A/H3N2 epidemic variants. Vaccine [Internet]. 2017;35(2):255–63. Available from: http://dx.doi.org/10.1016/j.vaccine.2016.11.084

Clarke SK, Cook GT, Egglestone SI, Hall TS, Miller DL, Reed SE, et al. A virus from epidemic

vomiting disease. Br. Med. J. 1972;3(July):86–9.

Colquhoun DR, Schwab KJ, Cole RN, Halden RU. Detection of norovirus capsid protein in authentic standards and in stool extracts by matrix-assisted laser desorption ionization and nanospray mass spectrometry. Appl Env. Microbiol [Internet]. 2006;72(4):2749–55. Available from: https://www.ncbi.nlm.nih.gov/pubmed/16597979

Compton J. Nucleic acid sequence-based amplification. Nature [Internet]. 1991;350(6313):91–2. Available from: http://www.nature.com/articles/350091a0

Corless CE, Guiver M, Borrow R, Edwards-Jones V, Kaczmarski EB, Fox AJ. Contamination and sensitivity issues with a real-time universal 16s rRNA PCR. J. Clin. Microbiol. 2000;38(5):1747–52.

Costantini V, Grenz L, Fritzinger A, Lewis D, Biggs C, Hale A, et al. Diagnostic accuracy and analytical sensitivity of IDEIA Norovirus assay for routine screening of human norovirus. J Clin Microbiol [Internet]. 2010;48(8):2770–8. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20554813

Coudray-Meunier C, Fraisse A, Martin-Latil S, Delannoy S, Fach P, Perelle S. A Novel High-Throughput Method for Molecular Detection of Human Pathogenic Viruses Using a Nanofluidic Real-Time PCR System. PLoS One [Internet]. 2016;11(1):e0147832. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26824897

Coudray-Meunier C, Fraisse A, Martin-Latil S, Guillier L, Delannoy S, Fach P, et al. A comparative study of digital RT-PCR and RT-qPCR for quantification of Hepatitis A virus and Norovirus in lettuce and water samples. Int J Food Microbiol [Internet]. 2015;201:17–26. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25725459

Coulson BS, Grimwood K, Hudson IL, Barnes GL, Bishop RF. Role of coproantibody in clinical protection of children during reinfection with rotavirus. J. Clin. Microbiol. 1992;30(7):1678–84.

Cox-Foster DL, Conlan S, Holmes EC, Palacios G, Evans JD, Moran NA, et al. A Metagenomic Survey of Microbes in Honey Bee Colony Collapse Disorder. Science (80-.). [Internet]. 2007;318(5848):283–7. Available from: http://www.sciencemag.org/cgi/doi/10.1126/science.1146498

Croci R, Tarantino D, Milani M, Pezzullo M, Rohayem J, Bolognesi M, et al. PPNDS inhibits murine Norovirus RNA-dependent RNA-polymerase mimicking two RNA stacking bases. FEBS Lett [Internet]. 2014;588(9):1720–5. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24657439

Croville G, Soubies SM, Barbieri J, Klopp C, Mariette J, Bouchez O, et al. Field monitoring of avian influenza viruses: Whole-genome sequencing and tracking of neuraminidase evolution using 454 pyrosequencing. J. Clin. Microbiol. 2012;50(9):2881–7.

Cubitt WD, Blacklow NR, Herrmann JE, Nowak N a, Nakata S, Chiba S. Antigenic relationships between human caliciviruses and Norwalk virus. J. Infect. Dis. 1987;156(5):806–14.

Cubitt WD, Green KY, Payment P. Prevalence of antibodies to the Hawaii strain of human calicivirus as measured by a recombinant protein based immunoassay. J. Med. Virol. 1998;54(2):135–9.

Czako R, Atmar RL, Opekun AR, Gilger MA, Graham DY, Estes MK. Serum hemagglutination inhibition activity correlates with protection from gastroenteritis in persons infected with Norwalk virus. Clin Vaccine Immunol [Internet]. 2012;19(2):284–7. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22190401

Damalanka VC, Kim Y, Alliston KR, Weerawarna PM, Galasiti Kankanamalage AC, Lushington GH, et al. Oxadiazole-Based Cell Permeable Macrocyclic Transition State Inhibitors of Norovirus 3CL Protease. J Med Chem [Internet]. 2016;59(5):1899–913. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26823007

Daughenbaugh KF, Fraser CS, Hershey JW, Hardy ME. The genome-linked protein VPg of the Norwalk virus binds eIF3, suggesting its role in translation initiation complex recruitment. EMBO J [Internet]. 2003;22(11):2852–9. Available from: https://www.ncbi.nlm.nih.gov/pubmed/12773399

Debbink K, Donaldson EF, Lindesmith LC, Baric RS. Genetic mapping of a highly variable norovirus GII.4 blockade epitope: potential role in escape from human herd immunity. J Virol [Internet]. 2012;86(2):1214–26. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22090110

Debbink K, Lindesmith LC, Ferris MT, Swanstrom J, Beltramello M, Corti D, et al. Within-host evolution results in antigenically distinct GII.4 noroviruses. J. Virol. [Internet]. 2014;88(13):7244–55. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4054459&tool=pmcentrez&rendertype=abstract

Delobel P, Saliou A, Nicot F, Dubois M, Trancart S, Tangre P, et al. Minor HIV-1 variants with the K103N resistance mutation during intermittent efavirenz-containing antiretroviral therapy and virological failure. PLoS One. 2011;6(6).

Dhanasekaran S, Doherty TM, Kenneth J. Comparison of different standards for real-time PCR-based absolute quantification. J. Immunol. Methods. 2010;354(1–2):34–9.

Dingle KE, Norovirus Infection Control in Oxfordshire Communities H. Mutation in a Lordsdale norovirus epidemic strain as a potential indicator of transmission routes. J Clin Microbiol [Internet]. 2004;42(9):3950–7. Available from: https://www.ncbi.nlm.nih.gov/pubmed/15364974

Dingle TC, Sedlak RH, Cook L, Jerome KR. Tolerance of droplet-digital PCR vs real-time quantitative PCR to inhibitory substances. Clin. Chem. 2013. p. 1670–2.

Division of Viral Diseases NC for I, Respiratory Diseases C for DC, Prevention. Updated norovirus outbreak management and disease prevention guidelines. MMWR Recomm Rep [Internet]. 2011;60(RR-3):1–18. Available from: https://www.ncbi.nlm.nih.gov/pubmed/21368741

Dohm JC, Lottaz C, Borodina T, Himmelbauer H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res. 2008;36(16).

Dolin R, Blacklow NR, DuPont H, Formal S, Buscho RF, Kasel J a, et al. Transmission of acute infectious nonbacterial gastroenteritis to volunteers by oral administration of stool filtrates. J. Infect. Dis. 1971;123(3):307–12.

Dolin R, Levy a G, Wyatt RG, Thornhill TS, Gardner JD. Viral gastroenteritis induced by the Hawaii agent. Jejunal histopathology and serologic response. Am. J. Med.

1975;59(December):761–8.

Dolin R, Reichman RC, Roessner KD, Tralka TS, Schooley RT, Gary W, et al. Detection by immune electron microscopy of the Snow Mountain agent of acute viral gastroenteritis. J. Infect. Dis. 1982;146(2):184–9.

Domingo E, Martinez-salas E, Sobrino F, Carlos J, Tortea D, Ortin J, et al. The quasispecies (extremely relevance - a review heterogeneous) nature of viral RNA genome populations : biological VandePol ', David Steinhauer ', Nicholas DePolo ' and John Holland ". 1985;40.

Domingo E, Menéndez-Arias L, Holland JJ. RNA virus fitness. Rev. Med. Virol. 1997. p. 87–96.

Domingo E, Sheldon J, Perales C. Viral Quasispecies Evolution. Microbiol. Mol. Biol. Rev. [Internet]. 2012;76(2):159–216. Available from: http://mmbr.asm.org/cgi/doi/10.1128/MMBR.05023-11

Dong HZQ, Wang G. Identification of Suitable Reference Genes for qRT-PCR Analysis of Circulating microRNAs in Hepatitis B Virus-Infected Patients. 2012;49–56.

Duizer E. Laboratory efforts to cultivate noroviruses. J. Gen. Virol. [Internet]. 2004 Jan 1 [cited 2014 Oct 23];85(1):79–87. Available from: http://vir.sgmjournals.org/cgi/doi/10.1099/vir.0.19478-0

Duizer E, van Duynhoven Y, Vennema H, Koopmans M. Failure to detect norovirus in a large group of asymptomatic individuals by Marshall et al. (Public Health Vol 118 (3) 230-233). Public Health [Internet]. 2004;118(6):455–7. Available from: https://www.ncbi.nlm.nih.gov/pubmed/15313600

Dunbar SA. Applications of Luminex(R) xMAPtrade mark technology for rapid, high-throughput multiplexed nucleic acid detection. Clin. Chim. Acta [Internet]. 2006;363(1–2):71–82. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16102740

Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, et al. Diversity of the human intestinal microbial flora. Science [Internet]. 2005;308(5728):1635–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15831718%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1395357

Eden JS, Tanaka MM, Boni MF, Rawlinson WD, White PA. Recombination within the pandemic norovirus GII.4 lineage. J Virol [Internet]. 2013;87(11):6270–82. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23536665

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res [Internet]. 2004;32(5):1792–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15034147

Eigen M. Selforganization of matter and the evolution of biological macromolecules. Naturwissenschaften. 1971;58(10):465–523.

Elftman MD, Gonzalez-Hernandez MB, Kamada N, Perkins C, Henderson KS, Nunez G, et al. Multiple effects of dendritic cell depletion on murine norovirus infection. J Gen Virol [Internet]. 2013;94(Pt 8):1761–8. Available from:

https://www.ncbi.nlm.nih.gov/pubmed/23636823

Eltahla AA, Lim KL, Eden JS, Kelly AG, Mackenzie JM, White PA. Nonnucleoside inhibitors of norovirus RNA polymerase: scaffolds for rational drug design. Antimicrob Agents Chemother [Internet]. 2014;58(6):3115–23. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24637690

Ettayebi K, Crawford SE, Murakami K, Broughman JR, Karandikar U, Tenge VR, et al. Replication of human noroviruses in stem cell-derived human enteroids. Science (80-.). [Internet]. 2016;353(6306):1387–93. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27562956

Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 1998;8(3):186–94.

Fankhauser RL, Monroe SS, Noel JS, Humphrey CD, Bresee JS, Parashar UD, et al. Epidemiologic and Molecular Trends of " Norwalk-like Viruses " Associated with Outbreaks of Gastroenteritis in the United States. 2000;1–7.

Faria NR, Azevedo R d. S d. S, Kraemer MUG, Souza R, Cunha MS, Hill SC, et al. Zika virus in the Americas: Early epidemiological and genetic findings. Science (80-.). [Internet]. 2016;352(6283):345–9. Available from: http://www.sciencemag.org/cgi/doi/10.1126/science.aaf5036

Fonager J, Stegger M, Rasmussen LD, Poulsen MW, Rønn J, Andersen PS, et al. A universal primer-independent next-generation sequencing approach for investigations of norovirus outbreaks and novel variants. Sci. Rep. [Internet]. 2017;7(1):813. Available from: http://www.nature.com/articles/s41598-017-00926-x

Fraisse A, Coudray-Meunier C, Martin-Latil S, Hennechart-Collette C, Delannoy S, Fach P, et al. Digital RT-PCR method for hepatitis A virus and norovirus quantification in soft berries. Int J Food Microbiol [Internet]. 2017;243:36–45. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27960104

Francis T. On the Doctrine of Original Antigenic Sin ON THE DOCTRINE OF ORIGINAL ANTIGENIC SIN *. Source Proc. Am. Philos. Soc. [Internet]. 1960;104(6):572–8. Available from: http://www.jstor.org/stable/985534%5Cnhttp://www.jstor.org/stable/985534?seq=1&cid=pdf-reference#references_tab_contents%5Cnhttp://about.jstor.org/terms

Franck KT, Fonager J, Ersboll AK, Bottiger B. Norovirus epidemiology in community and health care settings and association with patient age, Denmark. Emerg Infect Dis [Internet]. 2014;20(7):1123–31. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24960024

Frenck R, Bernstein DI, Xia M, Huang P, Zhong W, Parker S, et al. Predicting susceptibility to norovirus GII.4 by use of a challenge model involving humans. J Infect Dis [Internet]. 2012;206(9):1386–93. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22927452

Fu J, Li D, Xia S, Song H, Dong Z, Chen F, et al. Absolute quantification of plasmid DNA by real-time PCR with genomic DNA as external standard and its application to a biodistribution study of an HIV DNA vaccine. Anal. Sci. 2009;25(5):675–80.

Gairard-Dory AC, Degot T, Hirschi S, Schuller A, Leclercq A, Renaud-Picard B, et al. Clinical usefulness of oral immunoglobulins in lung transplant recipients with norovirus gastroenteritis: a case series. Transpl. Proc [Internet]. 2014;46(10):3603–5. Available from:

https://www.ncbi.nlm.nih.gov/pubmed/25498097

Galasiti Kankanamalage AC, Kim Y, Weerawarna PM, Uy RA, Damalanka VC, Mandadapu SR, et al. Structure-guided design and optimization of dipeptidyl inhibitors of norovirus 3CL protease. Structure-activity relationships and biochemical, X-ray crystallographic, cell-based, and in vivo studies. J Med Chem [Internet]. 2015;58(7):3144–55. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25761614

Gall A, Ferns B, Morris C, Watson S, Cotten M, Robinson M, et al. Universal amplification, next-generation sequencing, and assembly of HIV-1 genomes. J. Clin. Microbiol. 2012;50(12):3838–44.

Gall A, Kaye S, Hué S, Bonsall D, Rance R, Baillie GJ, et al. Restriction of V3 region sequence divergence in the HIV-1 envelope gene during antiretroviral treatment in a cohort of recent seroconverters. Retrovirology. 2013;10(1).

Gardy JL, Naus M, Amlani A, Chung W, Kim H, Tan M, et al. Whole-genome sequencing of measles virus genotypes H1 and D8 during outbreaks of infection following the 2010 Olympic Winter Games reveals viral transmission routes. J. Infect. Dis. 2015;212(10):1574–8.

Gentry-Shields J, Jaykus LA. Comparison of process control viruses for use in extraction and detection of human norovirus from food matrices. Food Res. Int. 2015;77:320–5.

Gerondopoulos A, Jackson T, Monaghan P, Doyle N, Roberts LO. Murine norovirus-1 cell entry is mediated through a non-clathrin-, non-caveolae-, dynamin- and cholesterol-dependent pathway. J Gen Virol [Internet]. 2010;91(Pt 6):1428–38. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20147520

Glass PJ, White LJ, Ball JM, Leparc-Goffart I, Hardy ME, Estes MK. Norwalk virus open reading frame 3 encodes a minor structural protein. J. Virol. 2000;74(14):6581–91.

Glass PJ, Zeng CQ, Estes MK. Two nonoverlapping domains on the Norwalk virus open reading frame 3 (ORF3) protein are involved in the formation of the phosphorylated 35K protein and in ORF3-capsid protein interactions. J Virol. 2003;77(6):3569–77.

Glassing A, Dowd SE, Galandiuk S, Davis B, Chiodini RJ. Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. Gut Pathog. 2016;8(1).

Glenister DA, Salamon KE, Smith K, Beighton D, Keevil CW. Enhanced Growth of Complex Communities of Dental Plaque Bacteria in Mucin-Limited Continuous Culture. Microb. Ecol. Health Dis. [Internet]. 1988;1(1):31–8. Available from: http://informahealthcare.com/doi/abs/10.3109/08910608809140176

Goldsmith CS, Miller SE. Modern uses of electron microscopy for detection of viruses. Clin. Microbiol. Rev. 2009;22(4):552–63.

Goodfellow I. The genome-linked protein VPg of vertebrate viruses - A multifaceted protein. Curr. Opin. Virol. 2011. p. 355–62.

Goodfellow I, Chaudhry Y, Gioldasi I, Gerondopoulos A, Natoni A, Labrie L, et al. Calicivirus translation initiation requires an interaction between VPg and eIF4E. EMBO Rep. [Internet]. 2005;6(10):968–72. Available from: http://embor.embopress.org/cgi/doi/10.1038/sj.embor.7400510

Gordon, I., H. S. Ingraham and RFK. Transmission of epidemic gastroenteritis to human volunteers by oral administration of fecal filtrates. J. Exp. Med. 1947;86:409–422.

Gra PE, Peyton C, Wheeler C, Apple R, Higuchi R, Shah K V. Reproducibility of HPV 16 and HPV 18 v iral load quantitation using TaqMan real-time PCR assays. 2003;112:23–33.

de Graaf M, van Beek J, Koopmans MP. Human norovirus transmission and evolution in a changing world. Nat Rev Microbiol [Internet]. 2016;14(7):421–33. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27211790

Green J, Gallimore CI, Norcott JP, Lewis D, Brown DW. Broadly reactive reverse transcriptase polymerase chain reaction for the diagnosis of SRSV-associated gastroenteritis. J Med Virol [Internet]. 1995;47(4):392–8. Available from: https://www.ncbi.nlm.nih.gov/pubmed/8636708

Green J, Norcott JP, Lewis D, Arnold C, Brown DWG. Norwalk-like viruses: Demonstration of genomic diversity by polymerase chain reaction. J. Clin. Microbiol. 1993;31(11):3007–12.

Green J, Wright PA, Gallimore CI, Mitchell O, Morgan-Capner P, Brown DW. The role of environmental contamination with small round structured viruses in a hospital outbreak investigated by reverse-transcriptase polymerase chain reaction assay. J Hosp Infect [Internet]. 1998;39(1):39–45. Available from: https://www.ncbi.nlm.nih.gov/pubmed/9617683

Green KY. Caliciviridae: The noroviruses. Fields Virol. [Internet]. 2007;2(v. 1):3177. Available from: http://www.loc.gov/catdir/toc/ecip072/2006032230.html

Green KY, Ando T, Balayan MS, Berke T, Clarke IN, Estes MK, et al. Taxonomy of the caliciviruses. J. Infect. Dis. 2000;181 Suppl(figure 1):S322–30.

Green KY, Mory a, Fogg MH, Weisberg a, Belliot G, Wagner M, et al. Isolation of enzymatically active replication complexes from feline calicivirus-infected cells. J.Virol. [Internet]. 2002;76(17):8582–95. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12163578%5Cnhttp://www.ncbi.nlm.nih.gov/pmc/articles/PMC136418/pdf/0319.pdf

Greenberg HB, Kapikian AZ. Detection of Norwalk agent antibody and antigen by solid-phase radioimmunoassay and immune adherence hemagglutination assay. J Am Vet Med Assoc [Internet]. 1978;173(5 Pt 2):620–3. Available from: https://www.ncbi.nlm.nih.gov/pubmed/100482

Greenberg HB, Valdesuso J, Kapikian a. Z, Chanock RM, Wyatt RG, Szmuness W, et al. Prevalence of antibody to the Norwalk virus in various countries. Infect. Immun. 1979;26(1):270–3.

Le Guyader FS, Bon F, DeMedici D, Parnaudeau S, Bertone A, Crudeli S, et al. Detection of multiple noroviruses associated with an international gastroenteritis outbreak linked to oyster consumption. J Clin Microbiol [Internet]. 2006;44(11):3878–82. Available from: https://www.ncbi.nlm.nih.gov/pubmed/17088365

Haga K, Fujimoto A, Takai-Todaka R, Miki M, Doan YH, Murakami K, et al. Functional receptor molecules CD300lf and CD300ld within the CD300 family enable murine noroviruses to infect cells. Proc Natl Acad Sci U S A [Internet]. 2016;113(41):E6248–55. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27681626

Haill CF, Newell P, Ford C, Whitley M, Cox J, Wallis M, et al. Compartmentalization of wards to cohort symptomatic patients at the beginning and end of norovirus outbreaks. J. Hosp. Infect. [Internet]. Elsevier Ltd; 2012;82(1):30–5. Available from: http://dx.doi.org/10.1016/j.jhin.2012.05.015

Hale A, Mattick K, Lewis D, Estes M, Jiang X, Green J, et al. Distinct epidemiological patterns of Norwalk-like virus infection. J Med Virol [Internet]. 2000;62(1):99–103. Available from: https://www.ncbi.nlm.nih.gov/pubmed/10935995

Hall AJ, Eisenbart VG, Etingue AL, Gould LH, Lopman BA, Parashar UD. Epidemiology of foodborne norovirus outbreaks, United States, 2001-2008. Emerg Infect Dis [Internet]. 2012;18(10):1566–73. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23017158

Hall AJ, Lopman BA, Payne DC, Patel MM, Gastanaduy PA, Vinje J, et al. Norovirus disease in the United States. Emerg Infect Dis [Internet]. 2013;19(8):1198–205. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23876403

Hansen KD, Brenner SE, Dudoit S. Biases in Illumina transcriptome sequencing caused by random hexamer priming. Nucleic Acids Res. 2010;38(12).

Hartnell R, Lowther J, Avant J, Dancer D, Lees D, Russell J. The development of LENTICULES as reference materials for noroviruses. J Appl Microbiol [Internet]. 2012;112(2):338–45. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22093409

Hasing ME, Hazes B, Lee BE, Preiksaitis JK, Pang XL. A next generation sequencing-based method to study the intra-host genetic diversity of norovirus in patients with acute and chronic infection. BMC Genomics [Internet]. 2016;17:480. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27363999

Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. Genomics. 2016. p. 1–8.

Heid Stevens, J., Livak, K.J., and Williams, P.M. C a. Real time quantitative PCR. Genome Res. 1996;6:986–94.

Henn MR, Boutwell CL, Charlebois P, Lennon NJ, Power KA, Macalalad AR, et al. Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. PLoS Pathog. 2012;8(3).

Henry S, Oriol R, Samuelsson B. Lewis Histo???Blood Group System and Associated Secretory Phenotypes. Vox Sang. 1995. p. 166–82.

Herbert TP, Brierley I, Brown TDK. Identification of a protein linked to the genomic and subgenomic mRNAs of feline calicivirus and its role in translation. J. Gen. Virol. 1997;78(5):1033–40.

Hewitt J, Leonard M, Greening GE, Lewis GD. Influence of wastewater treatment process and the population size on human virus profiles in wastewater. Water Res [Internet]. 2011;45(18):6267–76. Available from: https://www.ncbi.nlm.nih.gov/pubmed/21962483

Hillier LW, Marth GT, Quinlan AR, Dooling D, Fewell G, Barnett D, et al. Whole-genome sequencing and variant discovery in C. elegans. Nat. Methods [Internet]. 2008;5(2):183–8. Available from: http://www.nature.com/doifinder/10.1038/nmeth.1179

Hoa Tran TN, Trainor E, Nakagomi T, Cunliffe NA, Nakagomi O. Molecular epidemiology of

noroviruses associated with acute sporadic gastroenteritis in children: global distribution of genogroups, genotypes and GII.4 variants. J Clin Virol [Internet]. 2013;56(3):185–93. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23218993

Hodges RG, Mccorkle LP, Badger GF, Curtiss C, Dingle JH, Jordan WS, et al. A study of illness in a geoup of cleveland families. 1956;64:349–56.

Van den Hoecke S, Verhelst J, Vuylsteke M, Saelens X. Analysis of the genetic diversity of influenza A viruses using next-generation DNA sequencing. BMC Genomics [Internet]. 2015;16(1):79. Available from: http://www.biomedcentral.com/1471-2164/16/79

Hoffmann B, Scheuch M, Höper D, Jungblut R, Holsteg M, Schirrmeier H, et al. Novel orthobunyavirus in cattle, Europe, 2011. Emerg. Infect. Dis. 2012a;18(3):469–72.

Hoffmann D, Hutzenthaler M, Seebach J, Panning M, Umgelter A, Menzel H, et al. Norovirus GII.4 and GII.7 capsid sequences undergo positive selection in chronically infected patients. Infect. Genet. Evol. [Internet]. Elsevier B.V.; 2012b;12(2):461–6. Available from: http://dx.doi.org/10.1016/j.meegid.2012.01.020

Hohne M, Niendorf S, Mas Marques A, Bock CT. Use of sequence analysis of the P2 domain for characterization of norovirus strains causing a large multistate outbreak of norovirus gastroenteritis in Germany 2012. Int J Med Microbiol [Internet]. 2015;305(7):612–8. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26341330

Holley RW, Madison JT, Zamir A. A new method for sequence determination of large oligonucleotides. Biochem. Biophys. Res. Commun. [Internet]. 1964;17(4):389–94. Available from: http://www.sciencedirect.com/science/article/pii/0006291X64900178

Hoorfar J, Cook N, Malorny B, Wagner M, De Medici D, Abdulmawjood A, et al. Diagnostic PCR: Making internal amplification control mandatory. Lett. Appl. Microbiol. 2004. p. 79–80.

Hu L, Crawford SE, Czako R, Cortes-Penfield NW, Smith DF, Le Pendu J, et al. Cell attachment protein VP8* of a human rotavirus specifically interacts with A-type histo-blood group antigen. Nature [Internet]. 2012;485(7397):256–9. Available from: http://www.nature.com/doifinder/10.1038/nature10996

Huang P, Farkas T, Zhong W, Tan M, Thornton S, Morrow AL, et al. Norovirus and histo-blood group antigens: demonstration of a wide spectrum of strain specificities and classification of two major binding groups among multiple binding patterns. J Virol [Internet]. 2005;79(11):6714–22. Available from: https://www.ncbi.nlm.nih.gov/pubmed/15890909

Huggett J, Dheda K, Bustin S, Zumla A. Real-time RT-PCR normalisation ; strategies and considerations. 2005;279–84.

Huhti L, Blazevic V, Nurminen K, Koho T, Hytonen VP, Vesikari T. A comparison of methods for purification and concentration of norovirus GII-4 capsid virus-like particles. Arch Virol [Internet]. 2010;155(11):1855–8. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20721592

Huhti L, Szakal ED, Puustinen L, Salminen M, Huhtala H, Valve O, et al. Norovirus GII-4 causes a more severe gastroenteritis than other noroviruses in young children. J Infect Dis [Internet]. 2011;203(10):1442–4. Available from: https://www.ncbi.nlm.nih.gov/pubmed/21415019

Hunkapiller T, Kaiser RJ, Koop BF, Hood L. Large-scale and automated DNA sequence determination. Science (80-.). 1991;254(5028):59–67.

Hutson AM, Atmar RL, Graham DY, Estes MK. Norwalk virus infection and disease is associated with ABO histo-blood group type. J Infect Dis [Internet]. 2002;185(9):1335–7. Available from: https://www.ncbi.nlm.nih.gov/pubmed/12001052

Hyde JL, Gillespie LK, Mackenzie JM. Mouse norovirus 1 utilizes the cytoskeleton network to establish localization of the replication complex proximal to the microtubule organizing center. J Virol [Internet]. 2012;86(8):4110–22. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22301146

Hyde JL, Sosnovtsev S V, Green KY, Wobus C, Virgin HW, Mackenzie JM. Mouse norovirus replication is associated with virus-induced vesicle clusters originating from membranes derived from the secretory pathway. J Virol [Internet]. 2009;83(19):9709–19. Available from: https://www.ncbi.nlm.nih.gov/pubmed/19587041

Illingworth E, Taborn E, Fielding D, Cheesbrough J, Diggle PJ, Orr D. Is closure of entire wards necessary to control norovirus outbreaks in hospital? Comparing the effectiveness of two infection control strategies. J Hosp Infect [Internet]. 2011;79(1):32–7. Available from: https://www.ncbi.nlm.nih.gov/pubmed/21684038

Inouye S, Yamashita K, Yamadera S, Yoshikawa M, Kato N, Okabe N. Surveillance of viral gastroenteritis in Japan: pediatric cases and outbreak incidents. J. Infect. Dis. 2000;181 Suppl(April):S270–4.

Iritani N, Seto T, Hattori H, Natori K, Takeda N, Kubo H, et al. Humoral immune responses against norovirus infections of children. J Med Virol [Internet]. 2007;79(8):1187–93. Available from: https://www.ncbi.nlm.nih.gov/pubmed/17597489

Iturriza-Gomara M, Green J, Brown DWG, Desselberger U, Gray JJ. Comparison of specific and random priming in the reverse transcriptase polymerase chain reaction for genotyping group A rotaviruses. 1999;78:93–103.

Iturriza-Gomara M, Xerry J, Gallimore CI, Dockery C, Gray J. Evaluation of the Loopamp (loop-mediated isothermal amplification) kit for detecting Norovirus RNA in faecal samples. J Clin Virol [Internet]. 2008;42(4):389–93. Available from: https://www.ncbi.nlm.nih.gov/pubmed/18394955

Jabara CB, Jones CD, Roach J, Anderson JA, Swanstrom R. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. Proc. Natl. Acad. Sci. [Internet]. 2011;108(50):20166–71. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.1110064108

Jabs WJ, Hennig H, Kittel M, Pethig K, Smets O, Bucsky P, et al. Normalized Quantification by Real-Time PCR of Epstein-Barr Virus Load in Patients at Risk for Posttransplant Lymphoproliferative Disorders. 2001;39(2):564–9.

JD PMJKB. Quantitative competitive polymerase chain reaction for accurate quantitation of HIV DNA and RNA species. Biotechniques. Informa Life Sciences Publishing Group; 1993 Jan 1;14(1):70-.

Jiang X, Cubitt D, Hu J, Dai X, Treanor J, David OM, et al. Development of an ELISA to detect MX virus , a human calicivirus in the Snow Mountain agent genogroup. 1995;(1995):2739–47.

Jiang XI, Graham DY, Wang K, Estest MK. Norwalk Virus Genome Cloning and Characterization. 1990;250(iii).

Jiang XI, Wang J, Graham DY, Estes MK. Detection of Norwalk Virus in Stool by Polymerase Chain Reaction. 1992a;30(10):2529–34.

Jiang XI, Wang MIN, Graham DY, Estes MK. Expression , Self-Assembly , and Antigenicity of the Norwalk Virus Capsid Protein. 1992b;66(11):6527–32.

Johnson JA, Parra GI, Levenson EA, Green KY. A large outbreak of acute gastroenteritis in Shippensburg, Pennsylvania, 1972 revisited: evidence for common source exposure to a recombinant GII.Pg/GII.3 norovirus. Epidemiol Infect [Internet]. 2017;1–6. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28294087

Johnson PC, Mathewson JJ, DuPont HL, Greenberg HB. Multiple-challenge study of host susceptibility to Norwalk gastroenteritis in US adults. J Infect Dis [Internet]. 1990;161(1):18–21. Available from: https://www.ncbi.nlm.nih.gov/pubmed/2153184

Jones MK, Watanabe M, Zhu S, Graves CL, Keyes LR, Grau KR, et al. Enteric bacteria promote human and mouse norovirus infection of B cells. Science (80-.). [Internet]. 2014;346(6210):755–9. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25378626

Jonges M, Meijer A, Fouchier RA, Koch G, Li J, Pan JC, et al. Guiding outbreak management by the use of influenza a(H7NX) virus sequence analysis. Eurosurveillance. 2013;18(16).

Jonges M, Welkers MRA, Jeeninga RE, Meijer A, Schneeberger P, Fouchier RAM, et al. Emergence of the Virulence-Associated PB2 E627K Substitution in a Fatal Human Case of Highly Pathogenic Avian Influenza Virus A(H7N7) Infection as Determined by Illumina Ultra-Deep Sequencing. J. Virol. [Internet]. 2014;88(3):1694–702. Available from: http://jvi.asm.org/cgi/doi/10.1128/JVI.02044-13

Kageyama T, Kojima S, Shinohara M, Uchida K, Fukushi S, Hoshino FB, et al. Broadly reactive and highly sensitive assay for Norwalk-like viruses based on real-time quantitative reverse transcription-PCR. J Clin Microbiol [Internet]. 2003;41(4):1548–57. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12682144%5Cn0299899393 on real time PCR.pdf

Kageyama T, Shinohara M, Uchida K, Fukushi S, Hoshino FB, Kojima S, et al. Coexistence of multiple genotypes, including newly identified genotypes, in outbreaks of gastroenteritis due to Norovirus in Japan. J. Clin. Microbiol. [Internet]. 2004;42(7):2988–95. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15243049%5Cnhttp://www.ncbi.nlm.nih.gov/pmc/articles/PMC446284/pdf/1664-03.pdf

Kapikian a Z, Wyatt RG, Dolin R, Thornhill TS, Kalica a R, Chanock RM. Visualization by immune electron microscopy of a 27-nm particle associated with acute infectious nonbacterial gastroenteritis. J. Virol. 1972;10(5):1075–81.

Kapikian AZ. The Discovery of the 27-nm Norwalk Virus : An Historic Perspective. J. Infect. Dis. 2000;

Kaplan JE, Feldman R, Campbell DS, Lookabaugh C, Gary GW. The frequency of a Norwalk-like pattern of illness in outbreaks of acute gastroenteritis. Am J Public Heal. [Internet]. 1982;72(12):1329–32. Available from: https://www.ncbi.nlm.nih.gov/pubmed/6291414

Karlsson NG, Nordman H, Karlsson H, Carlstedt I, Hansson GC. Glycosylation differences between pig gastric mucin populations: a comparative study of the neutral oligosaccharides using mass spectrometry. Biochem. J. 1997;326:911–7.

Karst SM, Wobus CE, Lay M, Davidson J, Virgin HW th. STAT1-dependent innate immunity to a Norwalk-like virus. Science (80-.). [Internet]. 2003;299(5612):1575–8. Available from: https://www.ncbi.nlm.nih.gov/pubmed/12624267

Kasianowicz JJ, Brandin E, Branton D, Deamer DW. Characterization of individual polynucleotide molecules using a membrane channel. Proc. Natl. Acad. Sci. U. S. A. [Internet]. 1996;93(24):13770–3. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8943010%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC19421

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647–9.

Kellogg DE, Sninsky JJ, Kwok S. Quantitation of HIV-1 proviral DNA relative to cellular DNA by the polymerase chain reaction. Anal. Biochem. 1990;189:202–8.

Kim D, Hofstaedter CE, Zhao C, Mattei L, Tanes C, Clarke E, et al. Optimizing methods and dodging pitfalls in microbiome research. Microbiome. 2017.

Kirby A, Ashton L, Hart IJ. Detection of norovirus infection in the hospital setting using vomit samples. J Clin Virol [Internet]. 2011;51(1):86–7. Available from: https://www.ncbi.nlm.nih.gov/pubmed/21376659

Kirby A, Dove W, Ashton L, Hopkins M, Cunliffe NA. Detection of norovirus in mouthwash samples from patients with acute gastroenteritis. J Clin Virol [Internet]. 2010;48(4):285–7. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20637439

Kirby A, Iturriza-Gomara M. Norovirus diagnostics: options, applications and interpretations. Expert Rev Anti Infect Ther [Internet]. 2012;10(4):423–33. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22512752

Kitamoto N, Tanaka T, Natori K, Takeda N, Nakata S, Jiang X, et al. Cross-reactivity among several recombinant calicivirus virus-like particles (VLPs) with monoclonal antibodies obtained from mice immunized orally with one type of VLP. J Clin Microbiol [Internet]. 2002;40(7):2459–65. Available from: https://www.ncbi.nlm.nih.gov/pubmed/12089262

Kreutz LC, Seal BS, Mengeling WL. Early interaction of feline calicivirus with cells in culture. Arch. Virol. 1994;136(1–2):19–34.

Kroneman A, Harris J, Vennema H, Duizer E, van Duynhoven Y, Gray J, et al. Data quality of 5 years of central norovirus outbreak reporting in the European Network for food-borne viruses. J Public Heal. [Internet]. 2008a;30(1):82–90. Available from: https://www.ncbi.nlm.nih.gov/pubmed/18089585

Kroneman A, Vega E, Vennema H, Vinjé J, White P a., Hansman G, et al. Proposal for a unified norovirus nomenclature and genotyping. Arch. Virol. 2013;158(10):2059–68.

Kroneman A, Verhoef L, Harris J, Vennema H, Duizer E, van Duynhoven Y, et al. Analysis of integrated virological and epidemiological reports of norovirus outbreaks collected within the Foodborne Viruses in Europe network from 1 July 2001 to 30 June 2006. J Clin Microbiol

[Internet]. 2008b;46(9):2959–65. Available from: https://www.ncbi.nlm.nih.gov/pubmed/18650354

Kuchipudi S V, Tellabati M, Nelli RK, White GA, Perez BB, Sebastian S, et al. 18S rRNA is a reliable normalisation gene for real time PCR based on influenza virus infected cells. Virol. J. [Internet]. Virology Journal; 2012;9(1):1. Available from: Virology Journal

Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. 2016;33(7):1870–4.

Kundu S, Lockwood J, Depledge DP, Chaudhry Y, Aston A, Rao K, et al. Next-generation whole genome sequencing identifies the direction of norovirus transmission in linked patients. Clin Infect Dis [Internet]. 2013;57(3):407–14. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23645848

Lam TT, Zhu H, Smith DK, Guan Y, Holmes EC, Pybus OG. The recombinant origin of emerging human norovirus GII.4/2008: intra-genotypic exchange of the capsid P2 domain. J Gen Virol [Internet]. 2012;93(Pt 4):817–22. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22238233

Lamhoujeb S, Charest H, Fliss I, Ngazoa S, Jean J. Real-time molecular beacon NASBA for rapid and sensitive detection of norovirus GII in clinical samples. Can J Microbiol [Internet]. 2009;55(12):1375–80. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20029529

Lanata CF, Fischer-Walker CL, Olascoaga AC, Torres CX, Aryee MJ, Black RE, et al. Global causes of diarrheal disease mortality in children. PLoS One [Internet]. 2013;8(9):e72788. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24023773%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3762858

Lappalainen S, Blazevic V, Malm M, Vesikari T. Rotavirus vaccination and infection induce VP6-specific IgA responses. J. Med. Virol. 2017;89(2):239–45.

Larder BA, Kohli A, Kellam P, Kemp SD, Kronick M, Henfrey RD. Quantitative detection of HIV-1 drug resistance mutations by automated DNA sequencing. Nature. 1993;365(6447):671–3.

Laurence M, Hatzis C, Brash DE. Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. PLoS One. 2014;9(5).

Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, et al. Assessing the performance of the Oxford Nanopore Technologies MinION. Biomol. Detect. Quantif. 2015;3:1–8.

Lee GQ, Swenson LC, Poon AFY, Martin JN, Hatano H, Deeks SG, et al. Prolonged and Substantial Discordance in Prevalence of Raltegravir-Resistant HIV-1 in Plasma versus PBMC Samples Revealed by 454 "Deep" Sequencing. PLoS One. 2012;7(9).

Lee N, Chan MC, Wong B, Choi KW, Sin W, Lui G, et al. Fecal viral concentration and diarrhea in norovirus gastroenteritis. Emerg Infect Dis [Internet]. 2007;13(9):1399–401. Available from: https://www.ncbi.nlm.nih.gov/pubmed/18252121

Lessler J, Riley S, Read JM, Wang S, Zhu H, Smith GJD, et al. Evidence for antigenic seniority in influenza A (H3N2) antibody responses in southern China. PLoS Pathog. 2012;8(7):26.

Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016;44(W1):W242–5.

Lew, J F, Kapkian, A Z, Jiang, X, Estes, M K, Green KY. Molecular Characterization and Expression of the Capsid Protein of a Norwalk-like Virus Recovered from a Desert Shield Troop with Gastroenteritis. Virology; 1994. p. 319–25.

Li D, Breiman A, le Pendu J, Uyttendaele M. Binding to histo-blood group antigen-expressing bacteria protects human norovirus from acute heat stress. Front Microbiol [Internet]. 2015;6:659. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26191052

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009a;25(16):2078–9.

Li X, Zhou R, Wang Y, Sheng H, Tian X, Li H, et al. Identification and characterization of a native epitope common to norovirus strains GII/4, GII/7 and GII/8. Virus Res [Internet]. 2009b;140(1–2):188–93. Available from: https://www.ncbi.nlm.nih.gov/pubmed/19121346

Lima LRP, Silva AP da, Schmidt-Chanasit J, Paula VS de. Diagnosis of human herpes virus 1 and 2 (HHV-1 and HHV-2): use of a synthetic standard curve for absolute quantification by real time polymerase chain reaction. Mem. Inst. Oswaldo Cruz [Internet]. 2017;112(3):220–3. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0074-02762017000300220&lng=en&nrm=iso&tlng=en

Lindesmith L, Moe C, Marionneau S, Ruvoen N, Jiang X, Lindblad L, et al. Human susceptibility and resistance to Norwalk virus infection. Nat. Med. [Internet]. 2003;9(5):548–53. Available from: http://www.nature.com/doifinder/10.1038/nm860

Lindesmith LC, Beltramello M, Donaldson EF, Corti D, Swanstrom J, Debbink K, et al. Immunogenetic mechanisms driving norovirus GII.4 antigenic variation. PLoS Pathog [Internet]. 2012a;8(5):e1002705. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22615565

Lindesmith LC, Debbink K, Swanstrom J, Vinje J, Costantini V, Baric RS, et al. Monoclonal antibody-based antigenic mapping of norovirus GII.4-2002. J Virol [Internet]. 2012b;86(2):873–83. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22090098

Lindesmith LC, Donaldson EF, Beltramello M, Pintus S, Corti D, Swanstrom J, et al. Particle conformation regulates antibody access to a conserved GII.4 norovirus blockade epitope. J Virol [Internet]. 2014;88(16):8826–42. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24872579

Lindesmith LC, Donaldson EF, Lobue AD, Cannon JL, Zheng D-P, Vinje J, et al. Mechanisms of GII.4 norovirus persistence in human populations. PLoS Med. [Internet]. 2008 Feb [cited 2014 Sep 3];5(2):e31. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2235898&tool=pmcentrez&rendertype=abstract

Lindesmith LC, Ferris MT, Mullan CW, Ferreira J, Debbink K, Swanstrom J, et al. Broad blockade antibody responses in human volunteers after immunization with a multivalent norovirus VLP candidate vaccine: immunological analyses from a phase I clinical trial. PLoS Med [Internet]. 2015;12(3):e1001807. Available from:

https://www.ncbi.nlm.nih.gov/pubmed/25803642

Lindsay L, Wolter J, De Coster I, Van Damme P, Verstraeten T. A decade of norovirus disease risk among older adults in upper-middle and high income countries: a systematic review. BMC Infect Dis [Internet]. 2015;15:425. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26467099

Liu J, Kibiki G, Maro V, Maro A, Kumburu H, Swai N, et al. Multiplex reverse transcription PCR Luminex assay for detection and quantitation of viral agents of gastroenteritis. J Clin Virol [Internet]. 2011;50(4):308–13. Available from: https://www.ncbi.nlm.nih.gov/pubmed/21256076

Liu P, Yuen Y, Hsiao HM, Jaykus LA, Moe C. Effectiveness of liquid soap and hand sanitizer against Norwalk virus on contaminated hands. Appl Env. Microbiol [Internet]. 2010;76(2):394–9. Available from: https://www.ncbi.nlm.nih.gov/pubmed/19933337

Lopman BA, Reacher MH, van Duijnhoven Y, Hanon F-X, Brown D, Koopmans M. Viral Gastroenteritis Outbreaks in Europe, 1995 - 2000. J. Infect. Dis. 2003. p. 90–6.

LWOFF A, TOURNIER P, HORNE R. SYSTEM OF VIRUSES. COLD SPRING Harb. Symp. Quant. Biol. 1 BUNGTOWN RD, PLAINVIEW, NY 11724: COLD SPRING HARBOR LAB PRESS; 1962;27:51-.

Lysen M, Thorhagen M, Brytting M, Hjertqvist M, Andersson Y, Hedlund KO. Genetic diversity among food-borne and waterborne norovirus strains causing outbreaks in Sweden. J Clin Microbiol [Internet]. 2009;47(8):2411–8. Available from: https://www.ncbi.nlm.nih.gov/pubmed/19494060

Maalouf H, Zakhour M, Le Pendu J, Le Saux JC, Atmar RL, Le Guyader FS. Distribution in tissue and seasonal variation of norovirus genogroup I and II ligands in oysters. Appl Env. Microbiol [Internet]. 2010;76(16):5621–30. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20562271

MacCannell T, Umscheid CA, Agarwal RK, Lee I, Kuntz G, Stevenson KB, et al. Guideline for the prevention and control of norovirus gastroenteritis outbreaks in healthcare settings. Infect Control Hosp Epidemiol [Internet]. 2011;32(10):939–69. Available from: https://www.ncbi.nlm.nih.gov/pubmed/21931246

Majid L, Zagorodnyaya T, Plant EP, Petrovskaya S, Bidzhieva B, Ye Z, et al. Deep sequencing for evaluation of genetic stability of influenza A/California/07/2009 (H1N1) vaccine viruses. PLoS One. 2015;10(9).

Makino A, Shimojima M, Miyazawa T, Kato K, Tohya Y, Akashi H. Junctional adhesion molecule 1 is a functional receptor for feline calicivirus. J. Virol. [Internet]. 2006;80(9):4482–90. Available from: http://jvi.asm.org/content/80/9/4482.short

Margulies M, Egholm M, Altman W, Attiya S, Bader J. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437(September):376–80.

Marionneau S, Ruvoën N, Le Moullac–Vaidye B, Clement M, Cailleau–Thomas A, Ruiz–Palacois G, et al. Norwalk virus binds to histo-blood group antigens present on gastroduodenal epithelial cells of secretor individuals. Gastroenterology [Internet]. 2002 Jun [cited 2014 Nov 6];122(7):1967–77. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0016508502000306

Massin P, van der Werf S, Naffakh N. Residue 627 of PB2 is a determinant of cold sensitivity in RNA replication of avian influenza viruses. J. Virol. [Internet]. 2001;75(11):5398–404. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=114948&tool=pmcentrez&rendertype=abstract

Mastrangelo E, Pezzullo M, Tarantino D, Petazzi R, Germani F, Kramer D, et al. Structure-based inhibition of Norovirus RNA-dependent RNA polymerases. J Mol Biol [Internet]. 2012;419(3–4):198–210. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22446684

Mathijs E, Stals A, Baert L, Botteldoorn N, Denayer S, Mauroy A, et al. A review of known and hypothetical transmission routes for noroviruses. Food Env. Virol [Internet]. 2012;4(4):131–52. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23412887

Matsushima Y, Shimizu T, Ishikawa M, Komane A, Okabe N, Ryo A, et al. Complete Genome Sequence of a Recombinant GII.P16-GII.4 Norovirus Detected in Kawasaki City, Japan, in 2016. Genome Announc [Internet]. 2016;4(5). Available from: https://www.ncbi.nlm.nih.gov/pubmed/27795262

Matthews JE, Dickey BW, Miller RD, Felzer JR, Dawson BP, Lee AS, et al. The epidemiology of published norovirus outbreaks: a review of risk factors associated with attack rate and genogroup. Epidemiol Infect [Internet]. 2012;140(7):1161–72. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22444943

Mattison K, Corneau N, Berg I, Bosch A, Duizer E, Gutierrez-Aguirre I, et al. Development and validation of a microarray for the confirmation and typing of norovirus RT-PCR products. J Virol Methods [Internet]. 2011;173(2):233–50. Available from: https://www.ncbi.nlm.nih.gov/pubmed/21345352

Mattison K, Shukla A, Cook A, Pollari F, Friendship R, Kelton D, et al. Human noroviruses in swine and cattle. Emerg Infect Dis [Internet]. 2007;13(8):1184–8. Available from: https://www.ncbi.nlm.nih.gov/pubmed/17953089

Maunula L, Miettinen IT, von Bonsdorff CH. Norovirus outbreaks from drinking water. Emerg Infect Dis [Internet]. 2005;11(11):1716–21. Available from: https://www.ncbi.nlm.nih.gov/pubmed/16318723

Maxam a M, Gilbert W. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U. S. A. [Internet]. 1977;74(2):560–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/265521%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC392330

Mayo MA. A summary of taxonomic changes recently approved by ICTV. Arch. Virol. 2002. p. 1655–6.

McKernan KJ, Peckham HE, Costa GL, McLaughlin SF, Fu Y, Tsung EF, et al. Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. Genome Res. 2009;19(9):1527–41.

Meakins SM, Adak GK, Lopman BA, O'Brien SJ. General outbreaks of infectious intestinal disease (IID) in hospitals, England and Wales, 1992-2000. J. Hosp. Infect. 2003;53(1):1–5.

Medvedev A, Viswanathan P, May J, Korba B. Regulation of human norovirus VPg nucleotidylylation by ProPol and nucleoside triphosphate binding by its amino terminal sequence in vitro. Virology [Internet]. 2017;503:37–45. Available from:

https://www.ncbi.nlm.nih.gov/pubmed/28110248

Mesquita JR, Costantini VP, Cannon JL, Lin SC, Nascimento MS, Vinje J. Presence of antibodies against genogroup VI norovirus in humans. Virol J [Internet]. 2013;10:176. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23735311

Metzker ML. Sequencing technologies - the next generation. Nat. Rev. Genet. [Internet]. Nature Publishing Group; 2010;11(1):31–46. Available from: http://dx.doi.org/10.1038/nrg2626

Michael G. Rossmann JEJ. ICOSAHEDRAL RNA VIRUS STRUCTURE. Annu. Rev. Biochem. 1989;58:533–73.

Miura T, Sano D, Suenaga A, Yoshimura T, Fuzawa M, Nakagomi T, et al. Histo-blood group antigen-like substances of human enteric bacteria as specific adsorbents for human noroviruses. J Virol [Internet]. 2013;87(17):9441–51. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23804639

Miyamoto C, Miyamoto N, Yamamoto H, Imai K, Shinomura Y. Detection of fecal interferon-induced transmembrane protein messenger RNA for colorectal cancer screening. Oncol. Lett. 2011;2:95–100.

Montoya V, Olmstead A, Tang P, Cook D, Janjua N, Grebely J, et al. Deep sequencing increases hepatitis C virus phylogenetic cluster detection compared to Sanger sequencing. Infect. Genet. Evol. 2016;43:329–37.

Morillo SG, Luchs A, Cilli A, Ribeiro CD, de Cássia Compagnoli Carmona R, do Carmo Sampaio Tavares Timenetsky M. Norovirus GII.Pe Genotype: Tracking a Foodborne Outbreak on a Cruise Ship Through Molecular Epidemiology, Brazil, 2014. Food Environ. Virol. 2017;9(2):142–8.

Mullis KB, Faloona F a. [21] Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol. 1987;155:335–50.

Mumphrey SM, Changotra H, Moore TN, Heimann-Nichols ER, Wobus CE, Reilly MJ, et al. Murine norovirus 1 infection is associated with histopathological changes in immunocompetent hosts, but clinical disease is prevented by STAT1-dependent interferon responses. J Virol [Internet]. 2007;81(7):3251–63. Available from: https://www.ncbi.nlm.nih.gov/pubmed/17229692

Murakami K, Kurihara C, Oka T, Shimoike T, Fujii Y, Takai-Todaka R, et al. Norovirus binding to intestinal epithelial cells is independent of histo-blood group antigens. PLoS One [Internet]. 2013;8(6):e66534. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23799113

Nakamura S, Yang CS, Sakon N, Ueda M, Tougan T, Yamashita A, et al. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. PLoS One. 2009;4(1).

Nakanishi K, Tsugawa T, Honma S, Nakata S, Tatsumi M, Yoto Y, et al. Detection of enteric viruses in rectal swabs from children with acute gastroenteritis attending the pediatric outpatient clinics in Sapporo, Japan. J Clin Virol [Internet]. 2009;46(1):94–7. Available from: https://www.ncbi.nlm.nih.gov/pubmed/19608459

Nasheri N, Petronella N, Ronholm J, Bidawid S, Corneau N. Characterization of the Genomic

Diversity of Norovirus in Linked Patients Using a Metagenomic Deep Sequencing Approach. Front Microbiol [Internet]. 2017;8:73. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28197136

Nasu A, Marusawa H, Ueda Y, Nishijima N, Takahashi K, Osaki Y, et al. Genetic heterogeneity of hepatitis C virus in association with antiviral therapy determined by ultra-deep sequencing. PLoS One. 2011;6(9).

Nayak MK, Balasubramanian G, Sahoo GC, Bhattacharya R, Vinje J, Kobayashi N, et al. Detection of a novel intergenogroup recombinant Norovirus from Kolkata, India. Virology [Internet]. 2008;377(1):117–23. Available from: https://www.ncbi.nlm.nih.gov/pubmed/18555887

Neverov A, Chumakov K. Massively parallel sequencing for monitoring genetic consistency and quality control of live viral vaccines. Proc. Natl. Acad. Sci. U. S. A. [Internet]. 2010;107(46):20063–8. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.1012537107%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/21041640%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2993378

Newman RM, Kuntzen T, Weiner B, Berical A, Charlebois P, Kuiken C, et al. Whole genome pyrosequencing of rare hepatitis c virus genotypes enhances subtype classification and identification of naturally occurring drug resistance variants. J. Infect. Dis. 2013;208(1):17–31.

Ngazoa ES, Fliss I, Jean J. Quantitative study of persistence of human norovirus genome in water using TaqMan real-time RT-PCR. J Appl Microbiol [Internet]. 2008;104(3):707–15. Available from: https://www.ncbi.nlm.nih.gov/pubmed/17953687

Nice TJ, Baldridge MT, McCune BT, Norman JM, Lazear HM, Artyomov M, et al. Interferon-lambda cures persistent murine norovirus infection in the absence of adaptive immunity. Science (80-.). [Internet]. 2015;347(6219):269–73. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25431489

Nicot F, Sauné K, Raymond S, Jeanne N, Carcenac R, Lefebvre C, et al. Minority resistant HIV-1 variants and the response to first-line NNRTI therapy. J. Clin. Virol. 2015;62:20–4.

Niendorf S, Jacobsen S, Faber M, Eis-Hubinger AM, Hofmann J, Zimmermann O, et al. Steep rise in norovirus cases and emergence of a new recombinant strain GII.P16-GII.2, Germany, winter 2016. Euro Surveill [Internet]. 2017;22(4). Available from: https://www.ncbi.nlm.nih.gov/pubmed/28181902

Nilsson M, Hedlund KO, Thorhagen M, Larson G, Johansen K, Ekspong A, et al. Evolution of human calicivirus RNA in vivo: accumulation of mutations in the protruding P2 domain of the capsid leads to structural changes and possibly a new phenotype. J Virol [Internet]. 2003;77(24):13117–24. Available from: https://www.ncbi.nlm.nih.gov/pubmed/14645568

Noel JS, Fankhauser RL, Ando T, Monroe SS, Glass RI. Identification of a distinct common strain of “Norwalk-like viruses” having a global distribution. J Infect Dis [Internet]. 1999;179(6):1334–44. Available from: https://www.ncbi.nlm.nih.gov/pubmed/10228052

Nolan T, Hands RE, Bustin SA. Quantification of mRNA using real-time RT-PCR. Nat. Protoc. [Internet]. 2006;1(3):1559–82. Available from: http://www.nature.com/doifinder/10.1038/nprot.2006.236

Notomi T, Okayama H, Masubuchi H, Yonekawa T, Watanabe K, Amino N, et al. Loop-mediated isothermal amplification. Nucleic Acids Res. 2000;28:E63.

Nyren P. Enzymatic method for continuous monitoring of DNA polymerase activity. Anal Biochem [Internet]. 1987;167(2):235–8. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=2831754

Okhuysen PC, Jiang X, Ye L, Johnson PC, Estes MK. Viral shedding and fecal IgA response after Norwalk virus infection. J Infect Dis [Internet]. 1995;171(3):566–9. Available from: https://www.ncbi.nlm.nih.gov/pubmed/7876602

Orchard RC, Wilen CB, Doench JG, Baldridge MT, McCune BT, Lee YC, et al. Discovery of a proteinaceous cellular receptor for a norovirus. Science (80-.). [Internet]. 2016;353(6302):933–6. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27540007

Padmanabhan R, Padmanabhan R, Wu R. Nucleotide sequence analysis of DNA. IX. Use of oligonucleotides of defined sequence as primers in DNA sequence analysis. Biochem. Biophys. Res. Commun. 1972;48(5):1295–302.

Palmer S, Kearney M, Maldarelli F, Halvas EK, Bixby CJ, Bazmi H, et al. Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. J. Clin. Microbiol. 2005;43(1):406–13.

Pang S, Koyanagi Y, Miles S, Wiley C, Vinters HV, Chen ISY. High levels of unintegrated HIV-1 DNA in brain tissue of AIDS dementiapatients. Nature. 1990;343:85–9.

Parashar UD, Dow L, Fankhauser RL, Humphrey CD, Miller J, Ando T, et al. An outbreak of viral gastroenteritis associated with consumption of sandwiches: implications for the control of transmission by food handlers. Epidemiol Infect [Internet]. 1998;121(3):615–21. Available from: https://www.ncbi.nlm.nih.gov/pubmed/10030711

Parra GI, Bok K, Taylor R, Haynes JR, Sosnovtsev S V, Richardson C, et al. Immunogenicity and specificity of norovirus Consensus GII.4 virus-like particles in monovalent and bivalent vaccine formulations. Vaccine [Internet]. 2012;30(24):3580–6. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22469864

Parrino TA, Schreiber DS, Trier JS, Kapikian AZ, Blacklow NR. Clinical immunity in acute gastroenteritis caused by Norwalk agent. N Engl J Med [Internet]. 1977;297(2):86–9. Available from: https://www.ncbi.nlm.nih.gov/pubmed/405590

Pasloske BL, Walkerpeach CR, Obermoeller RD, Winkler M, Dubois DB. Armored RNA technology for production of ribonuclease-resistant viral RNA controls and standards. J. Clin. Microbiol. 1998;36(12):3590–4.

Patel MM, Widdowson M-A, Glass RI, Akazawa K, Vinjé J, Parashar UD. Systematic literature review of role of noroviruses in sporadic gastroenteritis. Emerg. Infect. Dis. [Internet]. 2008 Aug [cited 2014 Oct 23];14(8):1224–31. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2600393&tool=pmcentrez&rendertype=abstract

Payne DC, Vinje J, Szilagyi PG, Edwards KM, Staat MA, Weinberg GA, et al. Norovirus and medically attended gastroenteritis in U.S. children. N Engl J Med [Internet]. 2013;368(12):1121–30. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23514289

Penedos AR, Myers R, Hadef B, Aladin F, Brown KE. Assessment of the utility of whole genome sequencing of measles virus in the characterisation of outbreaks. PLoS One. 2015;10(11).

Perales C, Mateo R, Mateu MG, Domingo E. Insights into RNA Virus Mutant Spectrum and Lethal Mutagenesis Events: Replicative Interference and Complementation by Multiple Point Mutants. J. Mol. Biol. 2007;369(4):985–1000.

Pérez-Losada M, Arenas M, Galán JC, Palero F, González-Candelas F. Recombination in viruses: Mechanisms, methods of study, and evolutionary consequences. Infect. Genet. Evol. 2015. p. 296–307.

Perry JW, Wobus CE. Endocytosis of murine norovirus 1 into murine macrophages is dependent on dynamin II and cholesterol. J Virol [Internet]. 2010;84(12):6163–76. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20375172

Phillips G, Lopman B, Tam CC, Iturriza-Gomara M, Brown D, Gray J. Diagnosing norovirus-associated infectious intestinal disease using viral load. BMC Infect Dis [Internet]. 2009;9:63. Available from: https://www.ncbi.nlm.nih.gov/pubmed/19442278

Polo D, Schaeffer J, Fournet N, Le Saux JC, Parnaudeau S, McLeod C, et al. Digital PCR for Quantifying Norovirus in Oysters Implicated in Outbreaks, France. Emerg Infect Dis [Internet]. 2016;22(12):2189–91. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27869597

Polz MF, Cavanaugh CM. Bias in Template-to-Product Ratios in Multitemplate PCR. 1998;64(10):3724–30.

Prasad B V, Hardy ME, Dokland T, Bella J, Rossmann MG, Estes MK. X-ray crystallographic structure of the Norwalk virus capsid. Science (80-.). [Internet]. 1999;286(October):287–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10514371

Prasad B V, Rothnagel R, Jiang X, Estes MK. Three-dimensional structure of baculovirus-expressed Norwalk virus capsids. J. Virol. [Internet]. 1994a;68(8):5117–25. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8035511

Prasad BVV, Matson DO, Smith AW. Three-dimensional Structure of Calicivirus [Internet]. J. Mol. Biol. 1994b. p. 256–64. Available from: http://www.sciencedirect.com/science/article/pii/S0022283684714392

Pringle CR. Virus Taxonomy - San Diego 1998. Arch. Virol. 1999;143(7):1449–59.

Probert CSJ, Jones PRH, Ratcliffe NM. A novel method for rapidly diagnosing the causes of diarrhoea. Gut. 2004;53(1):58–61.

Puustinen L, Blazevic V, Huhti L, Szakal ED, Halkosalo A, Salminen M, et al. Norovirus genotypes in endemic acute gastroenteritis of infants and children in Finland between 1994 and 2007. Epidemiol Infect [Internet]. 2012;140(2):268–75. Available from: https://www.ncbi.nlm.nih.gov/pubmed/21489338

Qiao N, Wang S-M, Wang J-X, Kang B, Zhen S-S, Zhang X-J, et al. Variation analysis of norovirus among children with diarrhea in rural Hebei Province, north of China. Infect. Genet. Evol. [Internet]. Elsevier B.V.; 2017;53:199–205. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1567134817301946

Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature [Internet]. 2016;530(7589):228–32. Available from: http://www.nature.com/doifinder/10.1038/nature16996

Radford AD, Chapman D, Dixon L, Chantrey J, Darby AC, Hall N. Application of next-generation sequencing technologies in virology. J. Gen. Virol. 2012;93(PART 9):1853–68.

Rahamat-Langendoen JC, Lokate M, Scholvinck EH, Friedrich AW, Niesters HG. Rapid detection of a norovirus pseudo-outbreak by using real-time sequence based information. J Clin Virol [Internet]. 2013;58(1):245–8. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23880160

Ramani S, Estes MK, Atmar RL. Correlates of Protection against Norovirus Infection and Disease-Where Are We Now, Where Do We Go? PLoS Pathog [Internet]. 2016;12(4):e1005334. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27115358

Ramani S, Neill FH, Opekun AR, Gilger MA, Graham DY, Estes MK, et al. Mucosal and Cellular Immune Responses to Norwalk Virus. J Infect Dis [Internet]. 2015;212(3):397–405. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25635121

Ravn V, Dabelsteen E. Tissue distribution of histo-blood group antigens. APMIS. 2000;108(1):1–28.

Redd AD, Collinson-Streng AN, Chatziandreou N, Mullis CE, Laeyendecker O, Martens C, et al. Previously transmitted HIV-1 strains are preferentially selected during subsequent sexual transmissions. J. Infect. Dis. 2012;206(9):1433–42.

Reeck A, Kavanagh O, Estes MK, Opekun AR, Gilger MA, Graham DY, et al. Serological correlate of protection against norovirus-induced gastroenteritis. J Infect Dis [Internet]. 2010;202(8):1212–8. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20815703

Rocha-Pereira J, Jochmans D, Debing Y, Verbeken E, Nascimento MS, Neyts J. The viral polymerase inhibitor 2'-C-methylcytidine inhibits Norwalk virus replication and protects against norovirus-induced diarrhea and mortality in a mouse model. J Virol [Internet]. 2013;87(21):11798–805. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23986582

Rocha-Pereira J, Jochmans D, Neyts J. Prophylactic treatment with the nucleoside analogue 2'-C-methylcytidine completely prevents transmission of norovirus. J Antimicrob Chemother [Internet]. 2015;70(1):190–7. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25228588

Rockx B, De Wit M, Vennema H, Vinje J, De Bruin E, Van Duynhoven Y, et al. Natural history of human calicivirus infection: a prospective cohort study. Clin Infect Dis [Internet]. 2002;35(3):246–53. Available from: https://www.ncbi.nlm.nih.gov/pubmed/12115089

Rohayem J, Munch J, Rethwilm A. Evidence of recombination in the norovirus capsid gene. J Virol [Internet]. 2005;79(8):4977–90. Available from: https://www.ncbi.nlm.nih.gov/pubmed/15795283

Rohayem J, Robel I, Jager K, Scheffler U, Rudolph W. Protein-primed and de novo initiation of RNA synthesis by norovirus 3Dpol. J Virol [Internet]. 2006;80(14):7060–9. Available from: https://www.ncbi.nlm.nih.gov/pubmed/16809311

Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature [Internet].

2011;475(7356):348–52. Available from:
http://www.nature.com/doifinder/10.1038/nature10242

de Rougemont a, Ruvoen-Clouet N, Simon B, Estienney M, Elie-Caille C, Aho S, et al. Qualitative and quantitative analysis of the binding of GII.4 norovirus variants onto human blood group antigens. J. Virol. [Internet]. 2011 May [cited 2014 Oct 26];85(9):4057–70. Available from:
http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3126233&tool=pmcentrez&rendertype=abstract

Ruis C, Roy S, Brown JR, Allen DJ, Goldstein RA, Breuer J. The emerging GII.P16-GII.4 Sydney 2012 norovirus lineage is circulating worldwide, arose by late-2014 and contains polymerase changes that may increase virus transmission. PLoS One. 2017;12(6):1–9.

Ruiz-Palacios GM, Cervantes LE, Ramos P, Chavez-Munguia B, Newburg DS. Campylobacter jejuni binds intestinal H(O) antigen (Fucα1, 2Galβ1, 4GlcNAc), and fucosyloligosaccharides of human milk inhibit its binding and infection. J. Biol. Chem. 2003;278(16):14112–20.

Sabria A, Pinto RM, Bosch A, Bartolome R, Cornejo T, Torner N, et al. Norovirus shedding among food and healthcare workers exposed to the virus in outbreak settings. J Clin Virol [Internet]. 2016;82:119–25. Available from:
https://www.ncbi.nlm.nih.gov/pubmed/27479175

Sachs AB, Sarnow P, Hentze MW. Starting at the beginning, middle, and end: Translation initiation in eukaryotes. Cell. 1997. p. 831–8.

Sahoo MK, Holubar M, Huang CH, Mohamed-Hadley A, Liu Y, Waggoner JJ, et al. Detection of emerging vaccine-related polioviruses by deep sequencing. J. Clin. Microbiol. 2017;55(7):2162–71.

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich H a, et al. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science. 1985;230:1350–4.

Sakon N, Yamazaki K, Nakata K, Kanbayashi D, Yoda T, Mantani M, et al. Impact of genotype-specific herd immunity on the circulatory dynamism of norovirus: a 10-year longitudinal study of viral acute gastroenteritis. J Infect Dis [Internet]. 2015;211(6):879–88. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25210139

Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 2014;12(1).

van der Sanden S, Pallansch MA, van de Kassteele J, El-Sayed N, Sutter RW, Koopmans M, et al. Shedding of Vaccine Viruses with Increased Antigenic and Genetic Divergence after Vaccination of Newborns with Monovalent Type 1 Oral Poliovirus Vaccine. J. Virol. [Internet]. 2009;83(17):8693–704. Available from:
http://jvi.asm.org/cgi/doi/10.1128/JVI.02388-08

Sanger F, Brownlee GG, Barrell BG. A two-dimensional fractionation procedure for radioactive nucleotides. J. Mol. Biol. 1965;13(2):IN1-IN4.

Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J. Mol. Biol. 1975;94(3):441–8.

Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. [Internet]. 1977;74(12):5463–7. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.74.12.5463

Sarcey E, Serres A, Tindy F, Chareyre A, Ng S, Nicolas M, et al. Quantifying low-frequency revertants in oral poliovirus vaccine using next generation sequencing. J. Virol. Methods. 2017;246:75–80.

Schrader C, Schielke A, Ellerbroek L, Johne R. PCR inhibitors - occurrence, properties and removal. J. Appl. Microbiol. 2012. p. 1014–26.

Schwarz TF, Leo O. Immune response to human papillomavirus after prophylactic vaccination with AS04-adjuvanted HPV-16/18 vaccine: Improving upon nature. Gynecol. Oncol. 2008;110(3 SUPPL.1).

Sedlak RH, Kuypers J, Jerome KR. A multiplexed droplet digital PCR assay performs better than qPCR on inhibition prone samples. Diagn. Microbiol. Infect. Dis. 2014;80(4):285–6.

Seitz SR, Leon JS, Schwab KJ, Lyon GM, Dowd M, McDaniels M, et al. Norovirus infectivity in humans and persistence in water. Appl Env. Microbiol [Internet]. 2011;77(19):6884–8. Available from: https://www.ncbi.nlm.nih.gov/pubmed/21856841

Siebenga JJ, Beersma MFC, Vennema H, van Biezen P, Hartwig NJ, Koopmans M. High prevalence of prolonged norovirus shedding and illness among hospitalized patients: a model for in vivo molecular evolution. J. Infect. Dis. [Internet]. 2008 Oct 1 [cited 2014 Oct 23];198(7):994–1001. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18774885

Siebenga JJ, Vennema H, Renckens B, de Bruin E, van der Veer B, Siezen RJ, et al. Epochal evolution of GGII.4 norovirus capsid proteins from 1995 to 2006. J. Virol. [Internet]. 2007 Sep [cited 2014 Oct 23];81(18):9932–41. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2045401&tool=pmcentrez&rendertype=abstract

Simmonds P. Recombination and selection in the evolution of picornaviruses and other Mammalian positive-stranded RNA viruses. J Virol [Internet]. 2006;80(22):11124–40. Available from: https://www.ncbi.nlm.nih.gov/pubmed/16956935

Simmonds P, Karakasiliotis I, Bailey D, Chaudhry Y, Evans DJ, Goodfellow IG. Bioinformatic and functional analysis of RNA secondary structure elements among different genera of human and animal caliciviruses. Nucleic Acids Res [Internet]. 2008;36(8):2530–46. Available from: https://www.ncbi.nlm.nih.gov/pubmed/18319285

Simmons K, Gambhir M, Leon J, Lopman B. Duration of immunity to norovirus gastroenteritis. Emerg Infect Dis [Internet]. 2013;19(8):1260–7. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23876612

Singh BK, Leuthold MM, Hansman GS. Human noroviruses' fondness for histo-blood group antigens. J Virol [Internet]. 2015;89(4):2024–40. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25428879

Sinha R, Stanley G, Gulati GS, Ezran C, Travaglini KJ, Wei E, et al. Index Switching Causes "Spreading-Of-Signal" Among Multiplexed Samples In Illumina HiSeq 4000 DNA Sequencing. bioRxiv [Internet]. 2017;125724. Available from: https://www.biorxiv.org/content/early/2017/04/09/125724

Springer GF, Williamson P, Brandes WC. BLOOD GROUP ACTIVITY OF GRAM-NEGATIVE BACTERIA. J. Exp. Med. [Internet]. 1961;113(6):1077–93. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2137423&tool=pmcentrez&rendertype=abstract

Ståhlberg A, Håkansson J, Xian X, Semb H, Kubista M. Properties of the Reverse Transcription Reaction in mRNA Quantification. Clin. Chem. 2004;50(3):509–15.

Stuart AD, Brown TDK. Entry of Feline Calicivirus Is Dependent on Clathrin-Mediated Endocytosis and Acidification in Endosomes. J. Virol. [Internet]. 2006;80(15):7500–9. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1563722/pdf/2452-05.pdf%5Cnhttp://jvi.asm.org/cgi/doi/10.1128/JVI.02452-05

Sukhrie FH, Siebenga JJ, Beersma MF, Koopmans M. Chronic shedders as reservoir for nosocomial transmission of norovirus. J Clin Microbiol [Internet]. 2010;48(11):4303–5. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20810762

Sukhrie FH, Teunis P, Vennema H, Bogerman J, van Marm S, Beersma MF, et al. P2 domain profiles and shedding dynamics in prospectively monitored norovirus outbreaks. J Clin Virol [Internet]. 2013;56(4):286–92. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23294532

Supadej K, Khamrin P, Kumthip K, Kochjan P, Yodmeeklin A, Ushijima H, et al. Wide variety of recombinant strains of norovirus GII in pediatric patients hospitalized with acute gastroenteritis in Thailand during 2005 to 2015. Infect. Genet. Evol. 2017;52:44–51.

Suzuki MT, Giovannoni SJ. Bias Caused by Template Annealing in the Amplification of Mixtures of 16S rRNA Genes by PCR †. 1996;62(2):625–30.

Tacket CO, Sztein MB, Losonsky GA, Wasserman SS, Estes MK. Humoral, mucosal, and cellular immune responses to oral Norwalk virus-like particles in volunteers. Clin Immunol [Internet]. 2003;108(3):241–7. Available from: https://www.ncbi.nlm.nih.gov/pubmed/14499247

Takahashi D, Kim Y, Lovell S, Prakash O, Groutas WC, Chang KO. Structural and inhibitor studies of norovirus 3C-like proteases. Virus Res [Internet]. 2013;178(2):437–44. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24055466

Tam CC, Rodrigues LC, Viviani L, Dodds JP, Evans MR, Hunter PR, et al. Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. Gut. 2012;61(1):69–77.

Tamminen K, Huhti L, Vesikari T, Blazevic V. Pre-existing immunity to norovirus GII-4 virus-like particles does not impair de novo immune responses to norovirus GII-12 genotype. Viral Immunol [Internet]. 2013;26(2):167–70. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23438469

Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. [Internet]. 1993;10(3):512–26. Available from: https://www.researchgate.net/publication/14861841_Tamura_K_Nei_M_Estimation_of_the_number_of_nucleotide_substitutions_in_the_control_region_of_mtDNA_in_humans_and_chimpanzees_Mol_Biol_Evol_10_512-526

Tamura M, Natori K, Kobayashi M, Miyamura T, Takeda N. Inhibition of attachment of

virions of Norwalk virus to mammalian cells by soluble histone molecules. Arch Virol [Internet]. 2003;148(9):1659–70. Available from: https://www.ncbi.nlm.nih.gov/pubmed/14505080

Tan M, Huang P, Meller J, Zhong W, Farkas T, Jiang X. Mutations within the P2 domain of norovirus capsid affect binding to human histo-blood group antigens: evidence for a binding pocket. J. Virol. 2003;77(23):12562–71.

Tan M, Jiang X. Association of histo-blood group antigens with susceptibility to norovirus infection may be strain-specific rather than genogroup dependent. J Infect Dis [Internet]. 2008;198(6):940–3. Available from: https://www.ncbi.nlm.nih.gov/pubmed/18721066

Tan M, Jiang X. Norovirus Gastroenteritis , Carbohydrate Receptors , and Animal Models. 2010;6(8).

Tan M, Xia M, Cao S, Huang P, Farkas T, Meller J, et al. Elucidation of strain-specific interaction of a GII-4 norovirus with HBGA receptors by site-directed mutagenesis study. Virology [Internet]. 2008;379(2):324–34. Available from: https://www.ncbi.nlm.nih.gov/pubmed/18692213

Tarantino D, Pezzullo M, Mastrangelo E, Croci R, Rohayem J, Robel I, et al. Naphthalene-sulfonate inhibitors of human norovirus RNA-dependent RNA-polymerase. Antivir. Res [Internet]. 2014;102:23–8. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24316032

Taube S, Perry JW, McGreevy E, Yetming K, Perkins C, Henderson K, et al. Murine noroviruses bind glycolipid and glycoprotein attachment receptors in a strain-dependent manner. J Virol [Internet]. 2012;86(10):5584–93. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22438544

Taube S, Perry JW, Yetming K, Patel SP, Auble H, Shu L, et al. Ganglioside-linked terminal sialic acid moieties on murine macrophages function as attachment receptors for murine noroviruses. J Virol [Internet]. 2009;83(9):4092–101. Available from: https://www.ncbi.nlm.nih.gov/pubmed/19244326

Teunis PFM, Moe CL, Liu P, Miller SE, Lindesmith L, Baric RS, et al. Norwalk Virus : How Infectious is It ? 2008;1476(April):1468–76.

Thorne LG, Goodfellow IG. Norovirus gene expression and replication. J Gen Virol [Internet]. 2014;95(Pt 2):278–91. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24243731

Thornhill TS, Wyatt RG, Kalica a R, Dolin R, Chanock RM, Kapikian a Z. Detection by immune electron microscopy of 26- to 27-nm viruslike particles associated with two family outbreaks of gastroenteritis. J. Infect. Dis. 1977;135(1):20–7.

Thornley CN, Hewitt J, Perumal L, Van Gessel SM, Wong J, David SA, et al. Multiple outbreaks of a novel norovirus GII.4 linked to an infected post-symptomatic food handler. Epidemiol Infect [Internet]. 2013;141(8):1585–97. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23388349

Thys K, Verhasselt P, Reumers J, Verbist BMP, Maes B, Aerssens J. Performance assessment of the Illumina massively parallel sequencing platform for deep sequencing analysis of viral minority variants. J. Virol. Methods. 2015;221:29–38.

Tian P, Brandl M, Mandrell R. Porcine gastric mucin binds to recombinant norovirus particles and competitively inhibits their binding to histo-blood group antigens and Caco-2 cells. Lett Appl Microbiol [Internet]. 2005;41(4):315–20. Available from: https://www.ncbi.nlm.nih.gov/pubmed/16162137

Tian P, Engelbrektson A, Mandrell R. Two-log increase in sensitivity for detection of norovirus in complex samples by concentration with porcine gastric mucin conjugated to magnetic beads. Appl Env. Microbiol [Internet]. 2008;74(14):4271–6. Available from: https://www.ncbi.nlm.nih.gov/pubmed/18515489

Tian P, Yang D, Jiang X, Zhong W, Cannon JL, Burkhardt W, et al. Specificity and kinetics of norovirus binding to magnetic bead-conjugated histo-blood group antigens. J. Appl. Microbiol. 2010;109(5):1753–62.

Töpfer A, Höper D, Blome S, Beer M, Beerenwinkel N, Ruggli N, et al. Sequencing approach to analyze the role of quasispecies for classical swine fever. Virology. 2013;438(1):14–9.

Treanor JJ, Jiang X, Madore HP, Estes MK. Subclass-specific serum antibody responses to recombinant Norwalk virus capsid antigen (rNV) in adults infected with Norwalk, Snow Mountain, or Hawaii virus. J. Clin. Microbiol. 1993;31(6):1630–4.

Trépanier P, Payment P, Trudel M. Concentration of human respiratory syncytial virus using ammonium sulfate, polyethylene glycol or hollow fiber ultrafiltration. J. Virol. Methods [Internet]. 1981;3(4):201–11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7328162

Tung G, Macinga D, Arbogast J, Jaykus LA. Efficacy of commonly used disinfectants for inactivation of human noroviruses and their surrogates. J Food Prot [Internet]. 2013;76(7):1210–7. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23834796

Turner P, Turner C, Jankhot A, Helen N, Lee SJ, Day NP, et al. A longitudinal study of streptococcus pneumoniae carriage in a cohort of infants and their mothers on the Thailand-Myanmar border. PLoS One. 2012;7(5).

US FDA. Food Code 2013 [Internet]. Drugs. 2013. Available from: http://www.fda.gov/downloads/Food/GuidanceRegulation/RetailFoodProtection/FoodCode/UCM374510.pdf%5Cnhttp://www.fda.gov/Food/GuidanceRegulation/RetailFoodProtection/FoodCode/ucm374275.htm

Vanregenmortel MH. Applying the species concept to plant viruses Brief. Arch. Virol. 1989;104:1–17.

Vega E, Barclay L, Gregoricus N, Shirley SH, Lee D, Vinje J. Genotypic and epidemiologic trends of norovirus outbreaks in the United States, 2009 to 2013. J Clin Microbiol [Internet]. 2014a;52(1):147–55. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24172151

Vega E, Donaldson E, Huynh J, Barclay L, Lopman B, Baric R, et al. RNA populations in immunocompromised patients as reservoirs for novel norovirus variants. J Virol [Internet]. 2014b;88(24):14184–96. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25275120

Vennema H, de Bruin E, Koopmans M. Rational optimization of generic primers used for Norwalk-like virus detection by reverse transcriptase polymerase chain reaction. J Clin Virol [Internet]. 2002;25(2):233–5. Available from: https://www.ncbi.nlm.nih.gov/pubmed/12367660

Verbist BMP, Thys K, Reumers J, Wetzels Y, Van Der Borght K, Talloen W, et al. VirVarSeq: A low-frequency virus variant detection pipeline for Illumina sequencing using adaptive base-calling accuracy filtering. Bioinformatics. 2015;31(1):94–101.

Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature [Internet]. 2006;439(7074):344–8. Available from: http://www.nature.com/doifinder/10.1038/nature04388

Vinje J. Advances in laboratory methods for detection and typing of norovirus. J Clin Microbiol [Internet]. 2015;53(2):373–81. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24989606

Vinjé J, Green J, Lewis DC, Gallimore CI, Brown DW, Koopmans MP. Genetic polymorphism across regions of the three open reading frames of "Norwalk-like viruses". Arch. Virol. 2000;145(2):223–41.

Vinje J, Koopmans MP. Molecular detection and epidemiology of small round-structured viruses in outbreaks of gastroenteritis in the Netherlands. J Infect Dis [Internet]. 1996;174(3):610–5. Available from: https://www.ncbi.nlm.nih.gov/pubmed/8769621

Voelkerding K V., Dames SA, Durtschi JD. Next-generation sequencing:from basic research to diagnostics. Clin. Chem. 2009. p. 641–58.

Vogelstein B, Kinzler KW. Digital PCR. Genetics. 1999;96:9236–41.

W.S. J, I. G, W.R D. A STUDY OF ILLNESS IN A GROUP OF CLEVELAND FAMILIES VII. TRANSMISSION OF ACUTE NON-BACTERIAL GASTROENTERITIS TO VOLUNTEERS: EVIDENCE FOR TWO DIFFERENT ETIOLOGIC AGENTS. J. Exp. Med. 1953;98(5):461–73.

Wagner A, Blackstone N, Cartwright P, Dick M, Misof B, Snow P, et al. Society of Systematic Biologists Surveys of Gene Families Using Polymerase Chain Reaction : PCR Selection and PCR Drift. 1994;43(2):250–61.

Wang a M, Doyle M V, Mark DF. Quantitation of mRNA by the polymerase chain reaction. Proc. Natl. Acad. Sci. U. S. A. 1989;86(December):9717–21.

Wang J, Jiang X, Madore HP, Gray J, Desselberger U, Ando T, et al. Sequence diversity of small, round-structured viruses in the Norwalk virus group. J. Virol. 1994;68(9):5982–90.

Wang X, Liu F, Jiang L, Bao Y, Xiao Y, Wang H. Use of chimeric influenza viruses as a novel internal control for diagnostic rRT-PCR assays. Appl. Microbiol. Biotechnol. 2016;100(4):1667–76.

Waters A, Coughlan S, Hall WW. Characterisation of a novel recombination event in the norovirus polymerase gene. Virology [Internet]. 2007;363(1):11–4. Available from: https://www.ncbi.nlm.nih.gov/pubmed/17434556

Watson SJ, Welkers MRA, Depledge DP, Coulter E, Breuer JM, de Jong MD, et al. Viral population analysis and minority-variant detection using short read next-generation sequencing. Philos. Trans. R. Soc. Lond. B. Biol. Sci. [Internet]. 2013;368(1614):20120205. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23382427%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3678329

Welniak LA, Blazar BR, Murphy WJ. Immunobiology of Allogeneic Hematopoietic Stem Cell Transplantation. Annu. Rev. Immunol. [Internet]. 2007;25(1):139–70. Available from: http://www.annualreviews.org/doi/10.1146/annurev.immunol.25.022106.141606

White LJ, Ball JM, Hardy ME, Tanaka TN, Kitamoto N, Estes MK. Attachment and entry of recombinant Norwalk virus capsids to cultured human and animal cell lines. J Virol [Internet]. 1996;70(10):6589–97. Available from: https://www.ncbi.nlm.nih.gov/pubmed/8794293

White PA. Evolution of norovirus. Clin Microbiol Infect [Internet]. 2014;20(8):741–5. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24980204

Widdowson MA, Rockx B, Schepp R, van der Poel WH, Vinje J, van Duynhoven YT, et al. Detection of serum antibodies to bovine norovirus in veterinarians and the general population in the Netherlands. J Med Virol [Internet]. 2005;76(1):119–28. Available from: https://www.ncbi.nlm.nih.gov/pubmed/15779045

Wilhelm BT, Landry JR. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. Methods. 2009. p. 249–57.

Wobus CE, Karst SM, Thackray LB, Chang KO, Sosnovtsev S V, Belliot G, et al. Replication of Norovirus in cell culture reveals a tropism for dendritic cells and macrophages. PLoS Biol [Internet]. 2004;2(12):e432. Available from: https://www.ncbi.nlm.nih.gov/pubmed/15562321

Wong TH, Dearlove BL, Hedge J, Giess AP, Piazza P, Trebes A, et al. Whole genome sequencing and de novo assembly identifies Sydney-like variant noroviruses and recombinants during the winter 2012/2013 outbreak in England. Virol J [Internet]. 2013;10:335. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24220146

Woodward JM, Gkrania-Klotsas E, Cordero-Ng AY, Aravinthan A, Bandoh BN, Liu H, et al. The Role Of Chronic Norovirus Infection In The Enteropathy Associated With Common Variable Immunodeficiency. Am. J. Gastroenterol. [Internet]. 2015;110(2):320–7. Available from: http://www.nature.com/doifinder/10.1038/ajg.2014.432

Wyatt RG, Dolin R, Blacklow NR, DuPont HL, Buscho RF, Thornhill TS, et al. Comparison of three agents of acute infectious nonbacterial gastroenteritis by cross-challenge in volunteers. J. Infect. Dis. 1974;129(6):709–14.

Xerry J, Gallimore CI, Cubitt D, Gray JJ. Tracking environmental norovirus contamination in a pediatric primary immunodeficiency unit. J Clin Microbiol [Internet]. 2010;48(7):2552–6. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20444966

Xerry J, Gallimore CI, Iturriza-Gomara M, Allen DJ, Gray JJ. Transmission events within outbreaks of gastroenteritis determined through analysis of nucleotide sequences of the P2 domain of genogroup II noroviruses. J Clin Microbiol [Internet]. 2008;46(3):947–53. Available from: https://www.ncbi.nlm.nih.gov/pubmed/18216210

Xerry J, Gallimore CI, Iturriza-Gomara M, Gray JJ. Tracking the transmission routes of genogroup II noroviruses in suspected food-borne or environmental outbreaks of gastroenteritis through sequence analysis of the P2 domain. J Med Virol [Internet]. 2009;81(7):1298–304. Available from: https://www.ncbi.nlm.nih.gov/pubmed/19475614

Yang Y, Xia M, Tan M, Huang P, Zhong W, Pang XL, et al. Genetic and phenotypic characterization of GII-4 noroviruses that circulated during 1987 to 2008. J Virol [Internet].

2010;84(18):9595–607. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20592096

Yoon S-W, Webby RJ, Webster RG. Evolution and ecology of influenza A viruses. Curr. Top. Microbiol. Immunol. [Internet]. 2014;385(January):359–75. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24990620

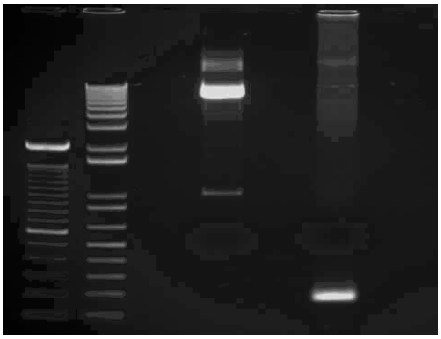
Zagordi O, Bhattacharya A, Eriksson N, Beerenwinkel N. ShoRAH: estimating the genetic diversity of a mixed sample from next-generation sequencing data. BMC Bioinformatics [Internet]. 2011;12(1):119. Available from: http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-119

Zahorsky J. Hyperemesis hiemis or the winter vomiting disease. Arch Pediatr [Internet]. 1929 Jan 1;46:391–5. Available from: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Hyperemesis+hiemis+or+the+winter+vomiting+disease#0

Zhang Z, Schwartz S, Wagner L, Miller W. A Greedy Algorithm for Aligning DNA Sequences. J. Comput. Biol. [Internet]. 2000;7(1–2):203–14. Available from: http://www.liebertonline.com/doi/abs/10.1089/10665270050081478

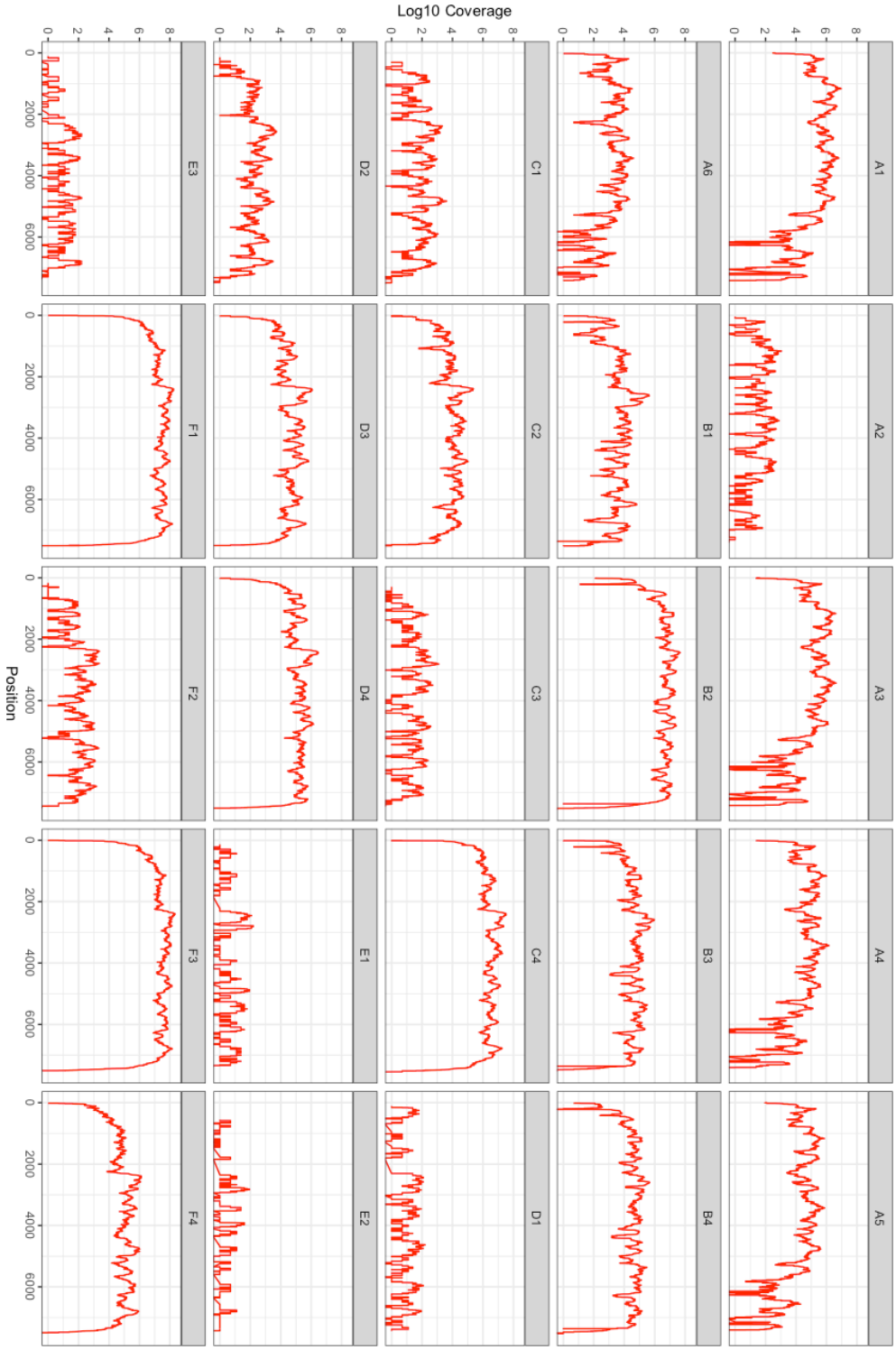
8 Appendix list

8.1 Appendix A – pCRGII3-3 (left)/RNA curve amplicon (right)

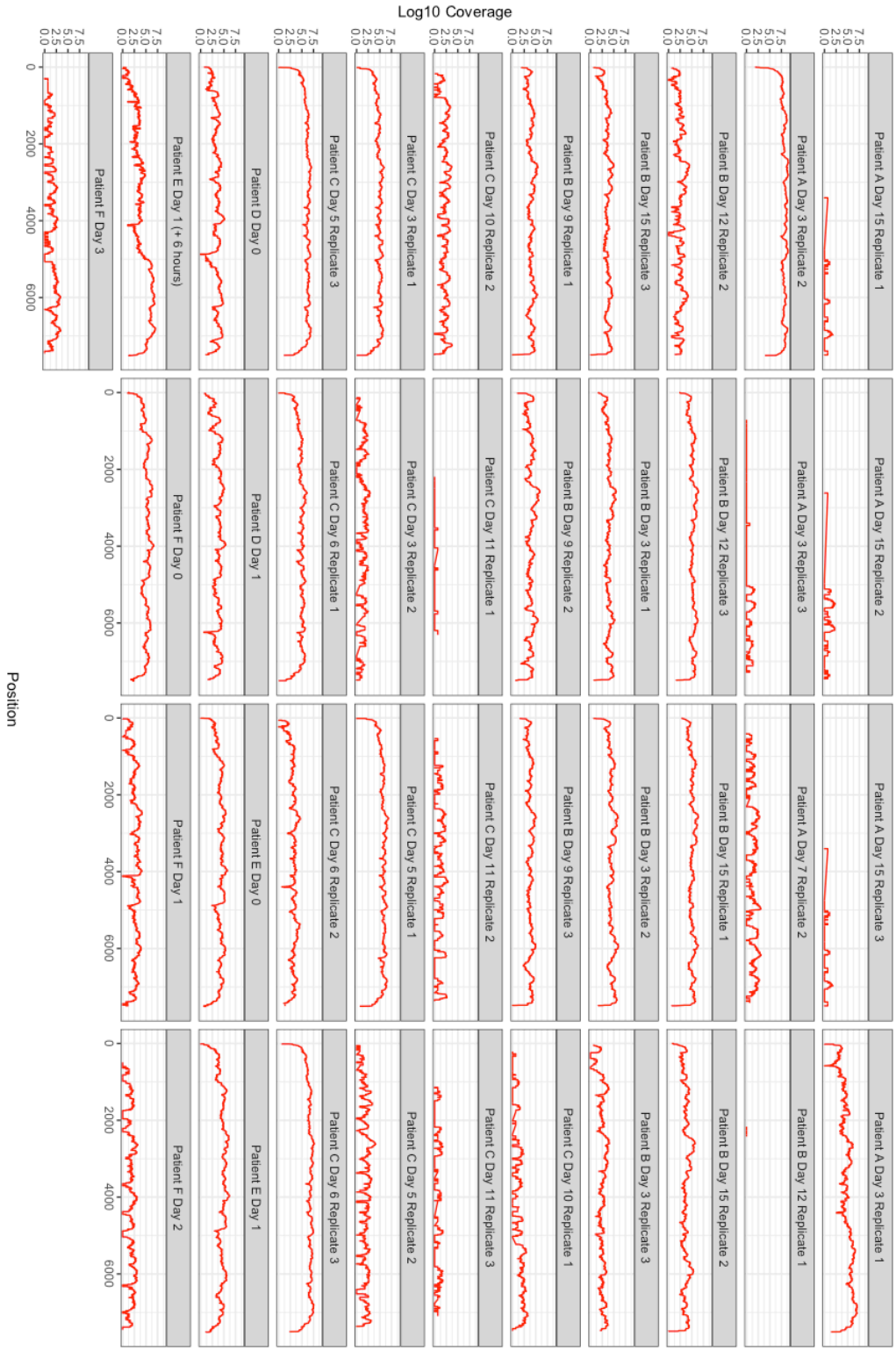


pCRGII3-3 (left)/RNA curve amplicon (right)

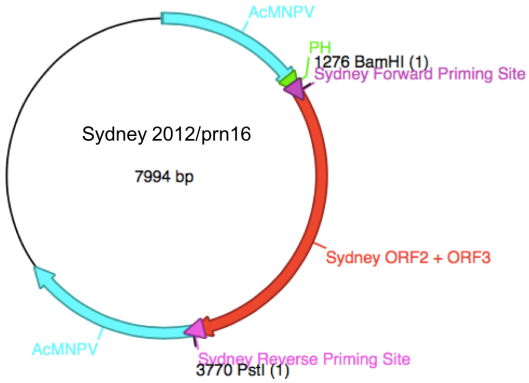
8.2 Appendix B – Transmission samples coverage graphs



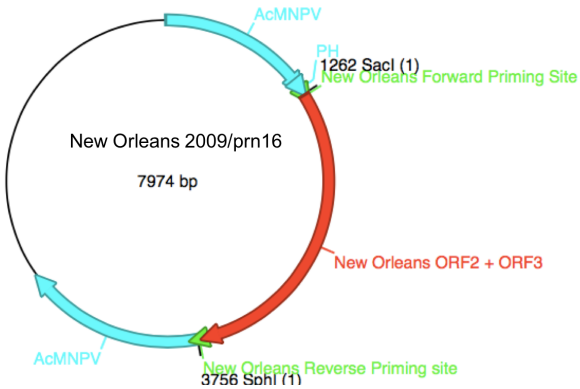
8.3 Appendix C – Longitudinal patient sample coverage graphs



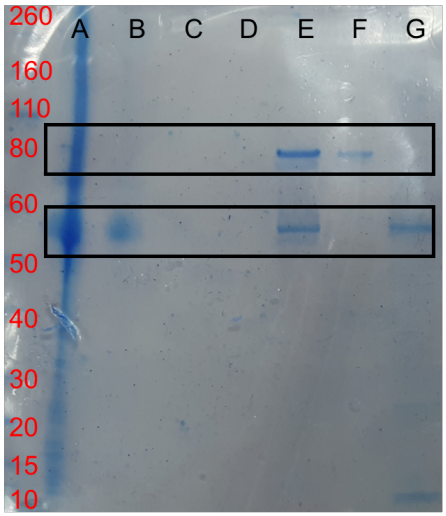
8.4 Appendix D – Sydney VLP plasmid



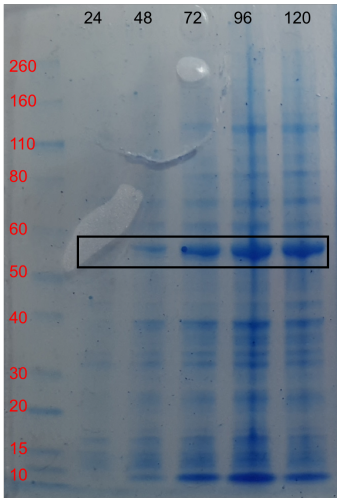
8.5 Appendix E – New Orleans VLP plasmid



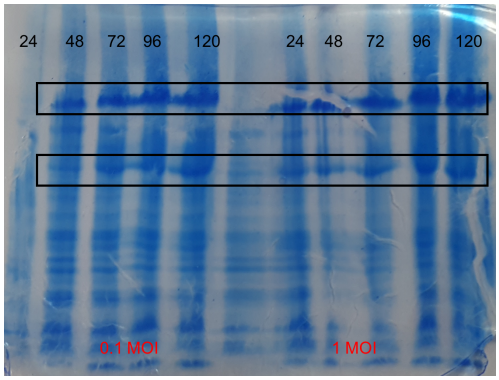
8.6 Appendix F – Farmington Hills post purification protein gel (A= cell lysate/E, F, G = sucrose cushion fractions)



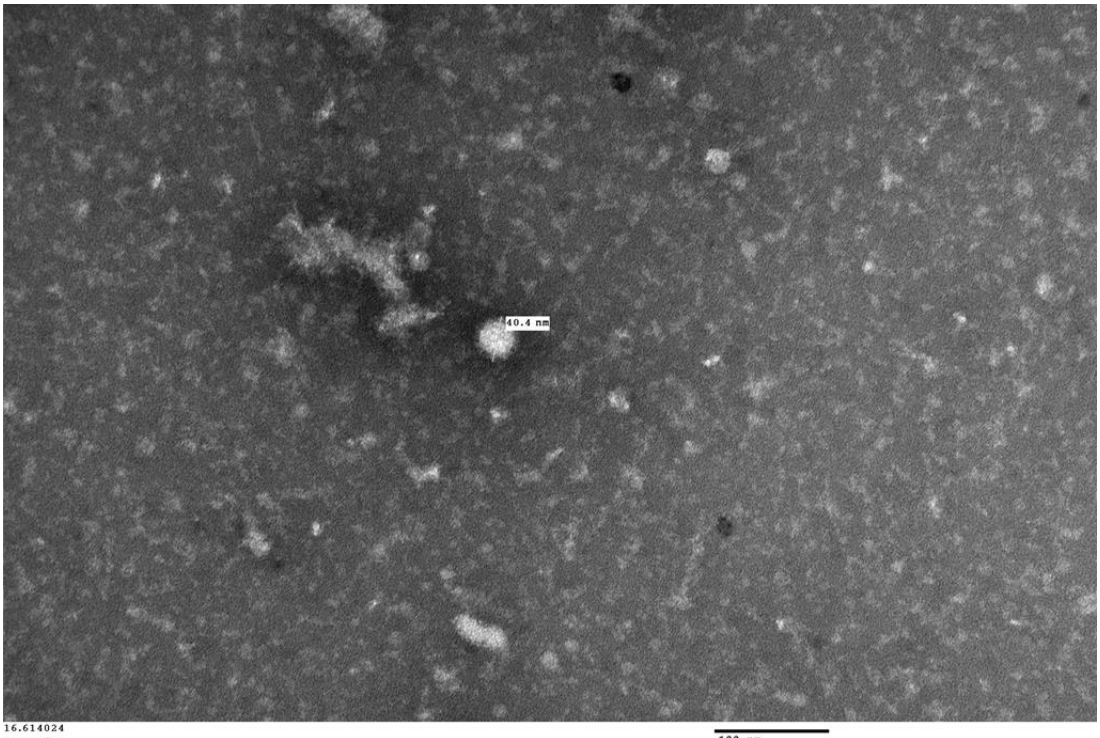
8.7 Appendix G – New Orleans recombinant baculovirus time course in hours (Black rectangle: The HuNoV major capsid protein)



8.8 Appendix H – Sydney recombinant baculovirus time course in hours (Black rectangles: The HuNoV major capsid protein and the aggregated form)



8.9 Appendix I Sydney – EM picture of Sydney strain VLPs



8.10 Appendix J – EM picture of New Orleans VLPs